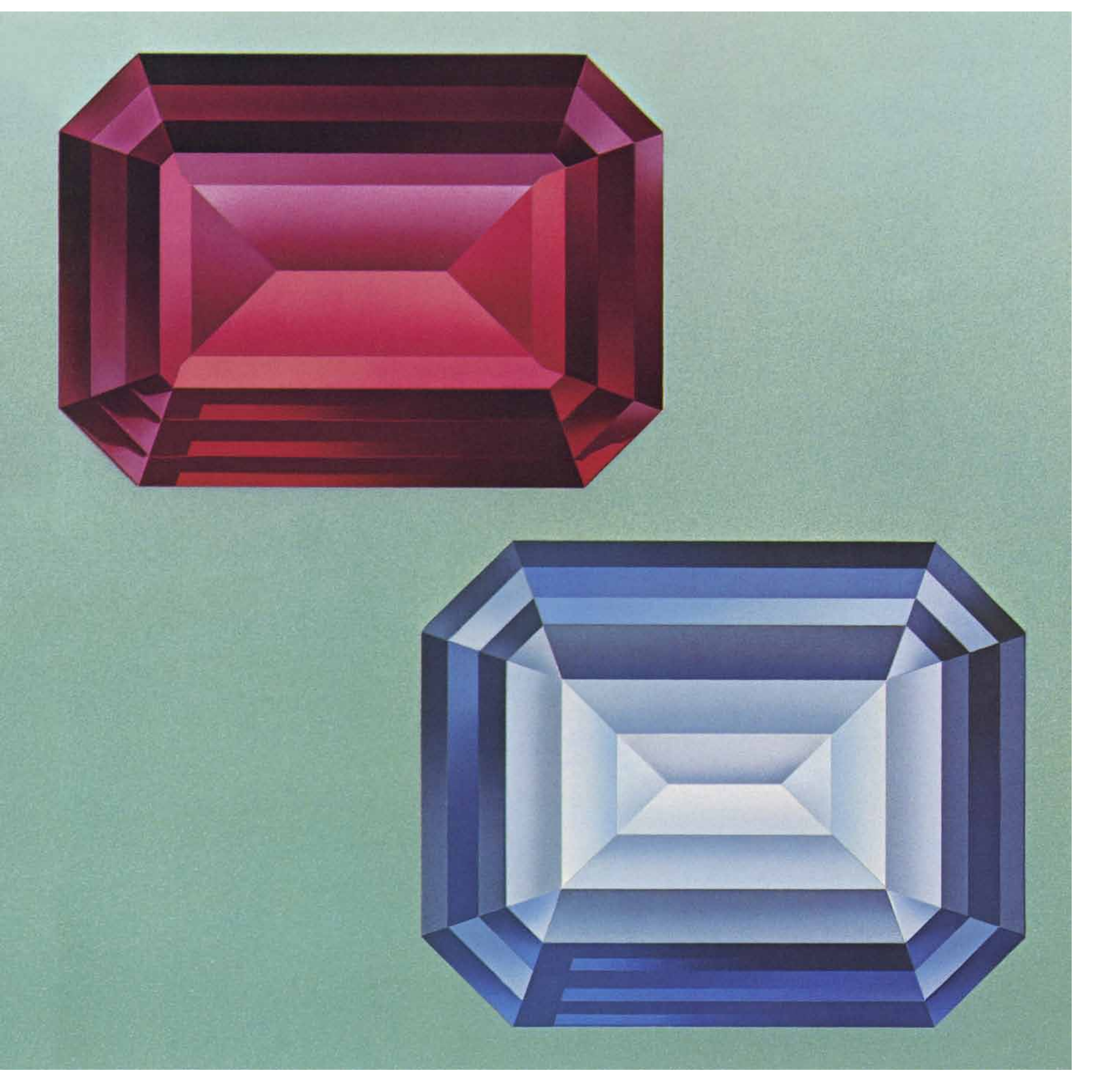
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



CAUSAS DEL COLOR

Diciembre 1980

250 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 51
Diciembre 1980

INVESTIGACION Y
CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

SCIENTIFIC AMERICAN

COMITE DE REDACCION Gerard Piel (Presidente), Dennis Flanagan, Brian P. Hayes, Philip Morrison, Francis Bello, Michael Feirtag, Judith Friedman, Paul W. Hoffman, Jonathan B. Piel, John Purcell, James T. Rogers, Armand Schwab, Jr., Joseph Wisnovsky

DIRECCION EDITORIAL Dennis Flanagan
DIRECCION ARTISTICA Samuel L. Howard
PRODUCCION Richard Sasso
DIRECTOR GENERAL George S. Conn

INVESTIGACION Y CIENCIA

DIRECTOR Francisco Gracia Guillén
REDACCION José María Valderas Gallardo (Redactor Jefe)
Carlos Oppenheimer
Encarna de Blas (Secretaria de Redacción)
PRODUCCION César Redondo Zayas
PROMOCION Y PUBLICIDAD Elena Sánchez-Fabres
PROMOCION EXTERIOR Pedro Clotas Cierco
EDITA Prensa Científica S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

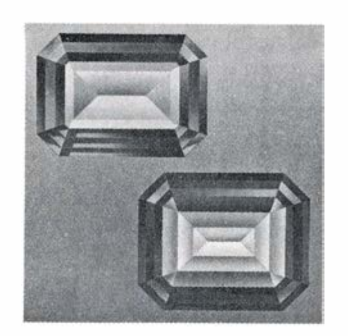
Colaboradores de este número:

Asesoramiento y traducción:

Manuel Puigcerver: *El curioso espectro de SS 433;* Romualdo Muñoz: *Anticuerpos monoclonales;* Amando García: *Las causas del color;* Montserrat Domingo: *Los Apalaches meridionales y el crecimiento de los continentes;* Miguel Aguilar: *Tomografía de emisión de positrones;* Juan Albaiges: *Minería del petróleo;* Joandomènec Ros: *La flotabilidad del nautilo;* Luis Bou: *Juegos matemáticos;* J. Vilardell: *Taller y laboratorio.*

Libros:

Ramón Margalef y Silverio Palafox



LA PORTADA

El cuadro de la portada representa dos piedras preciosas, un rubí y un zafiro azul, cuyas estructuras cristalinas básicas son idénticas, aunque el color de las mismas es radicalmente diferente. Ambas piedras están constituidas fundamentalmente por corindón, un óxido de aluminio que, en estado puro, es incoloro. Los colores resultan de la presencia de impurezas: cromo en el caso del rubí y titanio en el caso del zafiro azul. Los electrones de los átomos de la impureza absorben ciertas longitudes de onda de la luz, pero permiten que otras atraviesen estos cuerpos libremente y, por consiguiente, los cristales en cuestión son transparentes e intensamente coloreados. Los mecanismos según los cuales los electrones absorben ciertas longitudes de onda difieren en el rubí y el zafiro (véase "Las causas del color", de K. Nassau, pág. 56).

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2.750 pesetas
Extranjero:
Un año (12 números): 43 U.S.$
Ejemplar atrasado ordinario:
280 pesetas
Ejemplar atrasado extraordinario:
420 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49, Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 2.ª, Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Pintura de la portada de Ted Lodigensky

Página	Fuente
8-9	M. Llorens, A. Menzell y J. Villarrubia
10-20	Miguel Alonso
23	Eugene A. Harlan, Lick Observatory (*arriba*); C. Bruce Stephenson y Nicholas Sanduleak, Case Western Reserve University (*abajo*)
24	B. J. Geldzahler, MIT (*arriba*); Eric W. Greisen, National Radio Astronomy Observatory (*abajo*)
25	Frederick D. Seward, Centro Astrofísico del Observatorio de la Facultad de Harvard y del Observatorio Astrofísico Smithsoniano
26-33	Allen Beechel
39	Cesar Milstein y Georges Köhler
40-48	Bunji Tagawa
57	Fritz Goro
58-66	Dan Todd
67	Kurt Nassau
68	Dan Todd
70	Pierre Gilson
72	Ilil Arbel
75	Earth Satellite Corporation
76-77	Andrew Tomko
78	George V. Kelvin
80-81	George V. Kelvin; Frederick A. Cook, Larry D. Brown y Jack E. Oliver, Universidad de Cornell (*segundo empezando por arriba*)
82-86	George V. Kelvin
88	Michel M. Ter-Pogossian, Marcus E. Raichle y Burton E. Sobel, Facultad de Medicina de la Universidad de Washington
89	Ilil Arbel
90-93	Gabor Kiss
94	Michel M. Ter-Pogossian, Marcus E. Raichle y Burton E. Sobel, Facultad de Medicina de la Universidad de Washington
95-97	Gabor Kiss
98	Ralph Morse
99	Ilil Arbel
103-105	Walken Graphics
106	Syncrude Canada Ltd.
108	Walken Graphics
113	Peter Ward, Universidad de Califonia en Davis
114-117	Tom Prentiss
118	Lewis Greenwald, Universidad de Ohio
120	Peter Ward, Universidad de California en Davis
123-128	Ilil Arbel
131-135	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Nápoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

MARTIN LLORENS, ALFRED MENZELL y MIGUEL VILLARRUBIA ("Física y tecnología del reactor de fusión") han trabajado conjuntamente en la cátedra de física industrial de la Universidad de Barcelona en el campo de las técnicas energéticas bajo la dirección del profesor R. Marqués, recientemente fallecido. Llorens y Villarrubia son doctores en ciencias físicas por dicho centro, en donde, desde 1969, desempeñan tareas docentes y de investigación. Menzell es licenciado en ciencias físicas y trabaja en los laboratorios de la empresa Siemens en Munich; durante su permanencia en la cátedra de física industrial, realizó, bajo la dirección de Llorens, un interesante trabajo sobre la problemática del futuro reactor de fusión termonuclear.

BRUCE MARGON ("El curioso espectro de SS 433") ha ingresado recientemente en el departamento de astronomía de la Universidad de Washington. Se graduó en astrofísica por la Universidad de Columbia en 1968 y se doctoró por la de California en Berkeley en 1973. Tras un año como becario de la OTAN en el University College de Londres, regresó a Berkeley para incorporarse al equipo de investigación de la Universidad.

CESAR MILSTEIN ("Anticuerpos monoclonales") es miembro del British Medical Research Council Laboratory of Molecular Biology de Cambridge. Nacido en Argentina, estudió en la Universidad de Buenos Aires entre 1945 y 1952, graduándose en ciencias químicas. Permaneció en el instituto de química biológica de esa universidad, doctorándose en química en 1957. El año siguiente se incorporó al departamento de bioquímica de la Universidad de Cambridge. Tras dirigir la división de biología molecular del Instituto Microbiológico Nacional de Buenos Aires, Milstein regresó al laboratorio de biología molecular de Cambridge.

KURT NASSAU ("Las causas del color") está adscrito, como investigador, a los laboratorios Bell. Nacido en Austria, se trasladó a Gran Bretaña en 1938; estudió química y física en la Universidad de Bristol, por la que se graduó en 1948. A continuación marchó a los Estados Unidos, y trabajó como investigador químico, bioquímico y médico; en 1959 obtuvo el grado de doctor en química física por la Universidad de Pittsburgh. Desde entonces ha trabajado en los Laboratorios Bell, en el campo de la química de los cristales, especializándose en el estudio de los vidrios y cristales utilizados en la fabricación de láseres y fibras ópticas.

FREDERICK A. COOK, LARRY D. BROWN y JACK E. OLIVER ("Los Apalaches meridionales y el crecimiento de los continentes") se dedican a la geología y a la geofísica en la Universidad de Cornell. Cook se graduó por la Universidad de Wyoming en 1973. Allí completó sus estudios dos años después y trabajó como geofísico para la Continental Oil Company antes de trasladarse a Cornell. Brown ha sido profesor ayudante de ciencias geológicas en Cornell desde 1977. Oliver ha venido ocupando, desde 1971, la cátedra Irving Porter Church de ingeniería y la dirección del departamento de ciencias geológicas. Se formó en la Universidad de Columbia.

MICHEL M. TER-POGOSSIAN, MARCUS E. RAICHLE y BURTON E. SOBEL ("Tomografía de emisión de positrones") son profesores de la Facultad de Medicina de la Universidad de Washington. Ter-Pogossian estudió en el Lycée Michelet de París, licenciándose en matemática por la Universidad de París. Pasó un año en el Instituto del Radio, bajo la dirección de Irène Joliot-Curie, al tiempo que proseguía, en la universidad, sus trabajos de investigación en el campo de la física. Se doctoró en 1950 por la Universidad de Washington, coronando así sus trabajos en el campo de la física nuclear. Raichle es profesor de neurología y radiología. Después de doctorarse en medicina por la Universidad de Washington, se interesó por los temas de la circulación y el metabolismo cerebrales, siendo médico residente de neurología en el New York Hospital de la Facultad de Cornell. En 1971 se incorporó al claustro de la Facultad de Medicina de la Universidad de Washington, en la cual ha aplicado la tomografía de emisión de positrones a sus estudios sobre la circulación sanguínea y el metabolismo cerebrales. Sobel enseña medicina y dirige la división de cardiovascular. Se graduó por la Universidad de Cornell en 1958 y obtuvo el doctorado en medicina en 1962 por la Facultad de Medicina de Harvard. Tras varios años en el National Heart and Lung Institute, se incorporó a la Facultad de Medicina de la Universidad de California en San Diego. En 1973 Sobel se trasladó a la Universidad de Washington.

RICHARD A. DICK y SHELDON P. WIMPFEN ("Minería del petróleo") han colaborado durante muchos años como ingenieros con la industria minera y el Bureau of Mines de los Estados Unidos. Graduado en ingeniería minera por la Universidad Tecnológica de Michigan, Dick ha pertenecido al Bureau desde 1965. Wimpfen inició su carrera minera en 1934, como graduado, en la Escuela de Minas y Metalurgia de la Universidad de Texas en El Paso. Se incorporó al Bureau of Mines en 1970, siendo director adjunto de recursos minerales hasta 1975. Posteriormente aceptó el doble cargo de director adjunto de operaciones de campo e ingeniero jefe de minas, en los que se mantuvo hasta su jubilación en febrero último.

PETER WARD, LEWIS GREENWALD y OLIVE E. GREENWALD ("La flotabilidad del nautilo") comenzaron su colaboración en el estudio de este cefalópodo en la Universidad estatal de Ohio. Ward se formó en la Universidad de Washington y presentó su tesis doctoral en 1976 en la McMaster. Ha enseñado historia de la Tierra y paleobiología en la Estatal de Ohio desde 1976 a 1978; en la actualidad es profesor asistente de geología en la Universidad de California en Davis. "Me abrí camino en la universidad como buceador profesional de salvamento e instructor de escafandrismo", escribe. "Desde 1975, he pasado parte del año en el Indopacífico, estudiando el nautilo. Ello ha supuesto inmersiones nocturnas a gran profundidad, en las aguas exteriores de los taludes antearrecifales de Nueva Caledonia. Por lo que sé, soy el único investigador de cefalópodos que ha observado realmente al nautilo en su hábitat natural". Lewis y Olive Greenwald se incorporaron a la Universidad estatal de Ohio en 1972, donde en la actualidad son profesor asociado de zoología y profesor adjunto asistente de zoología, respectivamente. Lewis Greenwald se graduó en la Universidad de Syracuse, Olive Greenwald en la de Rice.

Hace...

José M.ª López Piñero

... doscientos años

Ignacio María Ruiz de Luzuriaga se trasladó a París con el fin de formarse en química, ciencias biológicas y medicina. Había nacido en la localidad vizcaína de Villaro diecisiete años antes. Su padre, José Santiago Ruiz de Luzuriaga, era un médico alavés amigo personal del Conde de Peñaflorida y miembro de la Sociedad Bascongada de Amigos del País, que había destacado por su labor como higienista, especialmente en la difusión de la variolización en el País Vasco.

Bajo la dirección de su padre, Ignacio María Ruiz de Luzuriaga había recibido una educación típicamente ilustrada. En 1777, cuando tenía quince años, ingresó en el Real Seminario Patriótico de Vergara, que acababa entonces de inaugurarse. Allí estudió humanidades, matemática, física y química. Entre sus maestros figuraron el físico Francisco Chabaneau y el gran químico Luis José Proust que, durante su primera estancia en España, dio un curso en el Seminario. Antes de abandonar Vergara, Luzuriaga presentó, como trabajo escolar, la traducción de un discurso sobre higiene naval que John Pringle había leído cuatro años antes en la Royal Society de Londres. Aparte de griego y de latín, su padre le había hecho estudiar inglés, francés e italiano.

Luzuriaga permaneció en París cuatro años, estudiando junto a varias celebridades científicas de la época, como los químicos Pierre Joseph Macquer y Antoine Fourcroy, los naturalistas Laurent de Jussieu y Louis Daubenton, el anatomista Antoine Portal y los médicos François Chopart y Pierre Desault. A finales de su estancia en la capital francesa, publicó su primer trabajo científico, relativo a la descomposición del aire atmosférico por el plomo, en la revista *Observations sur la physique, sur l'histoire naturelle et sur les arts,* dirigida por Jean Rozier. Fue también nombrado corresponsal del Jardín Botánico de Madrid y recibió una primera ayuda oficial, gestionada por el Conde de Aranda.

Desde 1785 a 1787, Luzuriaga estuvo en Gran Bretaña como "pensionado de S. M. para la Medicina y la Química", en las condiciones que reflejan los documentos y el epistolario publicados por Juan Riera. Tras pasar por Londres, marchó a Edimburgo, en cuya Universidad fue durante dos años discípulo, entre otros, del químico Joseph Black y del médico William Cullen. Obtuvo el título de maestro en artes y se doctoró en medicina con la tesis titulada *Tentamen medicum inaugurale de reciproca atque mutua systematis sanguinei et nervosi actione* (1786). Por otra parte, recorrió las manufacturas escocesas relacionadas con la química, entre otras, "la fábrica de cañones de Caran, la manufactura de vitriolo, la extracción de alquitrán del carbón de tierra y blanqueo de lienzos", así como varias minas. Fue nombrado miembro de las Reales Sociedades de Medicina e Historia Natural de Edimburgo.

Pasó a continuación una temporada en Glasgow, trabajando en las aplicaciones de la química a la medicina y a la biología. En 1787 asistió en Londres a los cursos del cirujano y naturalista John Hunter y frecuentó varios hospitales. Por último, volvió a pasar por París y trabajó algún tiempo en la Facultad de Medicina de Montpellier.

De regreso a España, tras una corta estancia en Bilbao junto a su padre, fijó su residencia en la Corte. Revalidó su título de médico, tras dos años reglamentarios de práctica, y permaneció en Madrid hasta su muerte en 1822. Su actividad profesional y científica estuvo principalmente ligada a la Real Academia Médica de Madrid.

En la obra de Luzuriaga pueden distinguirse dos aspectos fundamentales: sus investigaciones de laboratorio sobre cuestiones de química fisiológica y sus estudios acerca de diversas cuestiones de higiene pública.

Su primer trabajo de tema químico fue una memoria acerca de la descomposición del aire atmosférico por el plomo que, como hemos adelantado, publicó el año 1784 en las *Observations* de Rozier. Expuso en él una serie de experiencias con las que intentaba aclarar el proceso de lo que hoy llamamos oxidación del plomo y sus derivados, comparándolo con la de otros metales. La base doctrinal del estudio era todavía la teoría del flogisto de Stahl, dominante entonces en la comunidad científica europea. Continuó aceptándola en su tesis doctoral acerca de la interacción entre la sangre y el sistema nervioso (1786), en gran parte dedicada a investigar el mecanismo químico de la respiración mediante experimentos en animales y estudio de la acción *in vitro* de los gases sobre la sangre arterial y la venosa.

La culminación de este aspecto de la obra de Luzuriaga fue su *Disertación chimica fisiológica sobre la respiración y la sangre,* que presentó en abril de 1790 en la Real Academia Médica de Madrid y que más tarde apareció publicada en las *Memorias* de dicha institución. En ella se adhirió a la nueva teoría de Lavoisier, abandonando la del flogisto, desarrolló el contenido de su tesis doctoral y contestó a varias críticas que había recibido, en especial la que Louis Bernard Guyton de Morveau le había hecho desde las páginas de la *Encyclopédie methodique.*

En su célebre trabajo de 1780, Lavoisier y Laplace habían afirmado que la respiración era un proceso de oxidación localizado en los bronquios. La importancia de la contribución de Luzuriaga reside en que expuso por vez primera una teoría que situaba en la sangre circulante la combinación del oxígeno con el carbono. En la historia de la fisiología respiratoria fue un avance considerable en la línea que conduciría a la demostración por parte de Eduard Pflüger de que la respiración es un proceso intracelular (1872-1875). A finales del siglo XVIII, defendió una teoría parecida a la de Luzuriaga el científico francés Joseph Louis Lagrange, aunque no publicó nada sobre la materia, siendo conocidos sus puntos de vista por un trabajo (1791) de su discípulo Jean-Henri Hassenfratz. El historiador de la bioquímica Marcel Florkin opina que Luzuriaga y Lagrange llegaron independientemente a las mismas conclusiones. Manuel Usandizaga, en su libro sobre los Ruiz de Luzuriaga, destaca, entre las aportaciones de detalle contenidas en la *Disertación chimica fisiológica,* la demostración de que el cambio de color que sufre la sangre al pasar por el

pulmón se debe a la absorción de oxígeno. "La sangre venosa expuesta a la acción del gas oxígeno –afirma Luzuriaga– adquiere el color de la sangre arterial y se vicia este gas. Así el Doctor Grawford como Lavoisier coligen de esto que la mudanza de color que experimenta la sangre en la circulación procede de su combinación con el gas hidrógeno; y que al circular por el pulmón se desprende de su hidrógeno, tomando entonces su color encarnado o bermejo. De suerte que según estos sabios el gas oxígeno, que entra en la inspiración como parte constitutiva del aire atmosférico, se combina con el hidrógeno carbonoso que se está desprendiendo de la sangre, formando ácido carbónico con el carbono, y agua con el hidrógeno, poniéndose de color subido bermejo a proporción que pierde en la respiración el hidrógeno carbonoso de que estaba cargada. Pero voy a demostrar con experimentos directos que hay en la sangre un principio, bien sea el carbono, o bien el hidrógeno carbonoso, que combinándose con la base oxígena constituye el gas ácido carbónico, y que la mudanza de color que adquiere la sangre en los pulmones no procede tanto del desprendimiento de dicho principio, cuanto de la absorbencia del gas oxígeno." El último capítulo de la *Disertación,* dedicado a "las utilidades que resultan a la medicina de los puntos tratados en las secciones anteriores", destaca por ser uno de los más tempranos estudios que relaciona el oxígeno con las asfixias y con la reanimación de ahogados y sofocados.

La obra de Luzuriaga como higienista tuvo como supuestos los propios del "sanitary movement" de la Ilustración británica. En este campo, su primera publicación fue el *Tratado sobre el cólico de Madrid* (1797), consagrado a la investigación de las intoxicaciones, sobre todo por el plomo y el óxido de cobre, cuyo síntoma principal era el cólico, y a la forma de prevenirlas. Asocia el estudio químico, con el clínico y el medicosocial, desde una mentalidad que se refleja en la dedicatoria: "A la conservación de la salud pública. El Autor".

Aparte de otros trabajos de menor relieve, escribió más tarde memorias sobre la higiene de las cárceles (1802), la asistencia a los sordomudos (1817), la enseñanza clínica (1817) y la mortalidad en las inclusas (1817-1819). Esta última, titulada *Estadística Político Médico o estado comparativo de... Casas de Amparo u Hospicios de Maternidad, Ynclusas* y *Casas de Huerfanos...*, consta de cinco volúmenes. Carreras Panchón ha subrayado que fue una recapitulación de la política que el Estado ilustrado había seguido con los expósitos, utilizando datos cuantitativos como la mortalidad de las inclusas, los ingresos de las nodrizas y su influencia sobre la letalidad; ha hecho notar asimismo que una de sus lagunas es la edad de los niños muertos a partir del primer año.

Luzuriaga contribuyó también notablemente a la difusión de la vacunación antivariólica, así como a planificar la primera organización sanitaria española de carácter moderno.

TRATADO

SOBRE EL COLICO DE MADRID,

INSERTO EN LAS MEMORIAS

DE LA REAL ACADEMIA MEDICA DE MADRID;

Y PUBLICADO SEPARADAMENTE

DE ORDEN DE LA MISMA EN BENEFICIO COMUN:

POR

EL DOCTOR DON IGNACIO MARIA RUIZ de Luzuriaga, Socio de las Reales Sociedades de Medicina é Historia Natural de Edimburgo, del Liceo y Sociedad Médica de Londres, Académico de número de la Real Academia Médica Matritense, y Secretario del ramo de literatura y correspondencias extrangeras del Real Colegio Médico de esta Corte.

DE ORDEN SUPERIOR.

MADRID EN LA IMPRENTA REAL.

AÑO DE 1797.

Se hallará en la Librería de Don Antonio Baylo calle de las Carretas.

Portada del estudio que Luzuriaga dedicó a las intoxicaciones, sobre todo por el plomo y el óxido de cobre, cuyo síntoma principal era el cólico

Física y tecnología del reactor de fusión

Aunque el reactor de fusión termonuclear constituye la esperanza más sólida de obtención de energía a gran escala, los problemas físicos y tecnológicos que el mismo plantea son muchos y difíciles

Martín Llorens, Alfred Menzell y Miguel Villarrubia

Uno de los principales problemas con los que habrá de enfrentarse la humanidad en las próximas décadas será el del abastecimiento de energía. Suministro que se desea con las siguientes condiciones: supla la cada vez mayor escasez de reservas de combustibles, no contamine y cubra futuras demandas adicionales. Una solución muy prometedora a ese problema parece ser la energía de fusión termonuclear controlada, porque se trata de una fuente prácticamente inagotable, apenas contamina y es capaz de satisfacer nuevas demandas.

Las investigaciones sobre fusión termonuclear, que comenzaron hacia el año 1950 en varios países, han experimentado, tras un pequeño declive en la década de los sesenta, un nuevo auge en los últimos diez años. Con la próxima puesta en funcionamiento de grandes aparatos tokamak, que se hallan en fase de construcción en los Estados Unidos, Europa occidental, Rusia y Japón, y en los que se espera sostener la fusión termonuclear en un plasma de deuterio-tritio (D-T) durante intervalos de tiempo relativamente largos, las investigaciones sobre fusión controlada cobrarán un nuevo interés generalizado.

La fusión nuclear consiste en la unión de dos núcleos atómicos ligeros para formar otro núcleo más pesado y más estable, cuya masa es menor que la suma de las masas de los dos núcleos reaccionantes. La diferencia de masa se transforma en energía que se libera como energía cinética de los productos de reacción. Para la fusión controlada resultan de interés las reacciones siguientes: unión de dos núcleos de deuterio para formar un núcleo de helio 3 con una energía de 0,82 megaelectronvolt y un neutrón de una energía de 2,45 megaelectronvolt. A esta reacción la denominaremos reacción 1. Una segunda unión, a la que llamaremos reacción 2, se da entre dos núcleos de deuterio para formar un núcleo de tritio con una energía de 1,01 megaelectronvolt y un núcleo de hidrógeno de 3,02 megaelectronvolt. Una tercera unión (a la que daremos el nombre convencional de reacción 3) es la formada por un núcleo de deuterio y otro de tritio para constituir un núcleo de helio 4 con una energía de 3,52 megaelectronvolt y un neutrón de 14,06 megaelectronvolt. Por último, tenemos la unión de un núcleo de deuterio y otro de helio 3 para formar un núcleo de helio 4 con una energía de 3,6 megaelectronvolt y otro de hidrógeno de 14,7 megaelectronvolt. A esta última la llamaremos reacción 4.

El deuterio y el tritio son isótopos del hidrógeno. Las cuatro reacciones anteriores permiten varios ciclos de combustible. Tenemos así en primer lugar el ciclo deuterio-deuterio-tritio-helio 3 (que abreviaremos por su simbología D-D-T-^{3}He). Se basa en las cuatro reacciones o uniones antedichas. El helio 3 y el tritio que se producen por fusión de dos núcleos de deuterio (reacciones 1 y 2, respectivamente) se consumen en las dos reacciones restantes; la tercera, recordemos, se formaba por unión de un núcleo de deuterio con otro de tritio para constituir helio 4, y la cuarta, donde los núcleos reaccionantes eran el deuterio y el helio 3 para formar asimismo helio 4 e hidrógeno. En este ciclo, pues, los productos de reacción son, finalmente, núcleos de hidrógeno y helio y el combustible primario consumido es el deuterio. La energía que se libera aquí

CICLO D-D-T-^{3}He

D + D →	T	+	H	+	4,03 MeV
D + D →	^{3}He	+	n	+	3,27 MeV
T + D →	^{4}He	+	n	+	17,58 MeV
^{3}He + D →	^{4}He	+	H	+	18,30 MeV
6D →	2^{4}He	+	2H+2n	+	43,18 MeV

CICLO D-D-^{3}He

D + D →	T + H	+	4,03 MeV
D + D →	^{3}He + n	+	3,27 MeV
^{3}He + D →	^{4}He + H	+	18,30 MeV
5D →	^{4}He+T + 2H+n	+	25,60 MeV

CICLO D-T-^{6}Li

D + T →	^{4}He	+	n	+	17,58 MeV
^{6}Li + n →	^{4}He	+	T	+	4,78 MeV
D+^{6}Li →	2^{4}He			+	22,36 MeV

DISTINTOS CICLOS DE COMBUSTIBLES posibles. No es de prever que próximamente existan reactores que sigan los dos primeros y sí en cambio pronto podrían funcionar con el tercer ciclo D-T-Li.

PRINCIPALES CARACTERISTICAS DE ALGUNOS TOKAMAKS

Tokamak		Lugar	País	R (cm)	a (cm)	B_z (KG)	I_p (KA)	T_i (KeV)	T_e (KeV)	$n\tau$ (s.cm^{-3})
T4	(+)	Inst. Kurchatov	URSS	100	17	50	180	1	2	$3 \cdot 10^{11}$
T6	(+)	Inst. Kurchatov	URSS	70	22	15	150			
ST	(+)	Princeton	EEUU	109	13	40	130	0,5	2	$2 \cdot 10^{11}$
ATC	(+)	Princeton	EEUU	90-36	17-11	20-50	110-250			10^{11}
Alcator	(+)	MIT	EEUU	54	12	120	600	0,8	1	$1{,}2 \cdot 10^{13}$
Pulsator	(+)	Garching	RFA	70	11	28	100	0,3	0,8	$1{,}7 \cdot 10^{12}$
TPR	(+)	Fontenay aux R.	FRANCIA	98	20	60	400	1,8	3	$2 \cdot 10^{12}$
JFT2	(+)	Tokai-Mura	JAPON	90	25	10	180			
T10	(++)	Inst. Kurchatov	URSS	150	40	50	1000			
PLT	(++)	Princeton	EEUU	130	45	45	1400			
PDX	(++)	Princeton	EEUU	140	42	25	500			
ASDEX	(++)	Garching	RFA	165	40	28	500			
DITE	(++)	Culham	GB	112	23	30	280			
T20	(+++)		URSS	500	200	25	6000			
TFTR	(+++)	Princeton	EEUU	270	95	45	2500			
JET	(+++)	Culham	GB	296	125 horizontal 210 vertical	28	2600			
JT60	(+++)		JAPON	300	100	50	3000			

R = Radio mayor del toro a = Radio del plasma B_z = Campo magnético toroidal en el centro
I_p = Corriente en el plasma T_i, T_e = Temp. iones y electrones $n\tau$ = Parámetro de Lawson

(+) Tokamaks ensayados últimamente
(++) Tokamaks actualmente en funcionamiento
(+++) Tokamaks actualmente en proyecto o en fase de construcción

LAS ESPERANZAS DEPOSITADAS en los reactores de fusión se manifiestan en los numerosos dispositivos experimentales, del tipo tokamak, ensayados últimamente o en fase de proyecto o construcción en distintos países. Es posible que antes del año 2000 entren en funcionamiento algunos.

por unidad de masa de combustible (deuterio, repitamos) es 3,44 × 10^{14} joule por cada kilogramo, es decir, 3,98 megawatt-día por cada gramo. Un reactor de fusión que produjera 5000 megawatt térmicos de potencia consumiría 1,25 kilogramos de deuterio al día.

Un segundo ciclo de combustible está formado por la concatenación D-D-^{3}He. Este ciclo se funda en las reacciones primera, segunda y cuarta: la primera y la segunda partían de la unión de dos núcleos de deuterio y la cuarta de un núcleo de deuterio y otro de helio 3. El tritio generado en la segunda reacción, a partir de la fusión de dos núcleos de deuterio, no se consume ahora en la tercera reacción, aquella en que se fusionaban un núcleo de deuterio y otro de tritio para formar helio 4. ¿Qué ocurre pues con él? Se recupera de los productos de reacción que extraen del reactor las bombas de vacío. El tritio no vuelve a inyectarse, y se desintegra en helio 3 por emisión de rayos beta, con una vida media de 12,3 años. El combustible primario de este ciclo es también deuterio, y los productos de reacción son núcleos de hidrógeno, helio y tritio. La energía liberada por unidad de masa de combustible consumida alcanza los 2,45 × 10^{14} joule por cada kilogramo, es decir, 2,8 megawatt-día por cada gramo. El principal interés de este ciclo D-D-^{3}He reside en que casi un 90 por ciento de la energía liberada por fusión lo es como energía cinética de las partículas cargadas (partículas alfa, o núcleos de helio, y núcleos de tritio e hidrógeno) y, por tanto, puede utilizarse en un proceso de conversión directa a electricidad sin pasar por otras formas de energía. Sólo un 10 por ciento de la energía de fusión se libera como energía cinética de los neutrones, lo que implica un menor deterioro de los materiales de estructura del reactor por bombardeo neutrónico y, por consiguiente, se limita la generación de nuevos elementos radiactivos.

Ventaja de los dos ciclos de combustible mencionados es que las reservas de deuterio son prácticamente inagotables. Hallamos deuterio en el agua de los mares y océanos en una proporción del 0,016 por ciento. Y desventaja común de ambos es que las reacciones deuterio-deuterio en las dos primeras reacciones sólo se producen a temperaturas muy elevadas (del orden de los 500 millones de grados), lo que plantea graves dificultades físicas y técnicas.

Tenemos, por último, el ciclo deuterio-tritio-litio (abreviado D-T-Li), que se basa en la reacción tercera, aquella en que un núcleo de deuterio y otro de tritio se fusionaban para producir helio. Dicha reacción representa una sección eficaz (magnitud representativa de la frecuencia en que se produce) muy superior a las de las demás reacciones de fusión, y puede producirse además a temperaturas muy inferiores (del orden de los 50 millones de grados). Ello explica que sea ésta la reacción en que se fundarán los primeros reactores de fusión que entren en funcionamiento. En esta reacción de fusión se liberará una cantidad de energía del orden de los 17,58 megaelectronvolt, de los que un 20 por ciento será en forma de energía cinética de las partículas alfa (núcleos de helio) y el 80 por ciento restante como energía del neutrón.

El problema que presenta este último ciclo se centra en la producción artificial de tritio, que se consigue a través de las dos reacciones siguientes: primera, unión de un núcleo de litio 7 con un neutrón para formar un núcleo de helio, otro de tritio y un neutrón, con un déficit energético de 2,47 megaelectronvolt (reacción 5); y segunda, unión de un núcleo de litio 6 con un neutrón para formar un núcleo de helio y otro de tritio con una liberación de energía de 4,78 megaelectronvolt (reacción 6). En estas dos reacciones, los neutrones reaccionantes se aprovechan de los que se producen en la reacción, ya explicada, en que un núcleo de deuterio se une a otro de tritio para formar helio y neutrones (la tercera reacción antes expuesta).

Cuando se unen un núcleo de litio 7 y un neutrón para formar helio, tritio y neutrón, la reacción absorbe energía –es endotérmica–, se produce con neutrones rápidos y la magnitud representativa de la frecuencia con que ocurre es del orden de décimas de barn. (1 barn equivale a 10^{-24} centímetros cuadrados.) En esta reacción se libera un neutrón lento, que puede volver a reaccionar con el litio 6 para formar helio y tritio, según acabamos de exponer. Esta última reacción del litio 6 y el neutrón lento desprende energía –es exotérmica–, se produce con neutrones

lentos y tiene una sección eficaz del orden del barn. El litio 7 natural consta de un 92,58 por ciento de litio 7 y de un 7,42 de litio 6. Las reservas de litio existentes en la tierra se estiman suficientes para abastecernos de tritio durante centenares de miles de años.

A las elevadísimas temperaturas a las que deberá trabajarse para conseguir la fusión nuclear D-T, el deuterio y el tritio se encuentran en estado de plasma, es decir, constituyendo un gas de electrones y núcleos cargados positivamente, desligados entre sí. Para hacer posible la fusión nuclear controlada en un plasma D-T, deben cumplirse dos condiciones: en primer lugar, calentar el plasma hasta una temperatura del orden de los 50 millones de grados al objeto de dotar a los núcleos de D y de T de suficiente energía para que puedan chocar venciendo su propia repulsión electrostática o coulombiana (poseen carga del mismo signo) y fusionarse; y, en segundo lugar, hay que mantener confinado el plasma con una densidad suficiente y durante el tiempo necesario para que el producto de la densidad del plasma *n* por el tiempo de confinamiento τ sea mayor que $10^{14} - 10^{15}$ segundos por centímetro cúbico. El producto $n\tau$ se llama parámetro de Lawson. Esta segunda condición es necesaria para dar a los núcleos suficientes posibilidades de que entren en mutua colisión, de suerte que la energía generada por fusión en el plasma baste para compensar las pérdidas por radiación y por difusión de partículas hacia el exterior del confinamiento. En definitiva, para que el balance de energía sobre el plasma sea positivo.

Se trabaja hoy en dos formas de obtención de fusión nuclear controlada: el confinamiento inercial y el confinamiento magnético. Se funda el confinamiento inercial en el bombardeo de una gota sólida de deuterio-tritio con rayos láser desde varias direcciones, calentándola rápidamente y comprimiéndola hasta densidades muy altas. Y ello con el propósito de que, antes de que la gota se expanda a velocidad finita por su propia inercia, se libere suficiente energía de fusión para compensar la energía invertida en el calentamiento y compresión de la gota.

El confinamiento magnético se funda en el hecho de que las partículas cargadas no pueden moverse en dirección transversal al campo magnético, sino que se ven obligadas a describir trayectorias helicoidales a lo largo de las líneas de fuerza del campo magnético. Podemos confinar una masa de plasma porque la presión cinética del mismo puede compensarse con la presión que ejerce el campo magnético sobre el plasma, transversalmente a las líneas de fuerza. Esta presión del campo magnético se debe a que éste actúa sobre el plasma al interaccionar con corrientes que circulan en el interior del propio plasma, corrientes que vienen provocadas por la presencia del campo magnético aplicado; las fuerzas que desarrolla el campo magnético compensan las fuerzas de expansión del plasma.

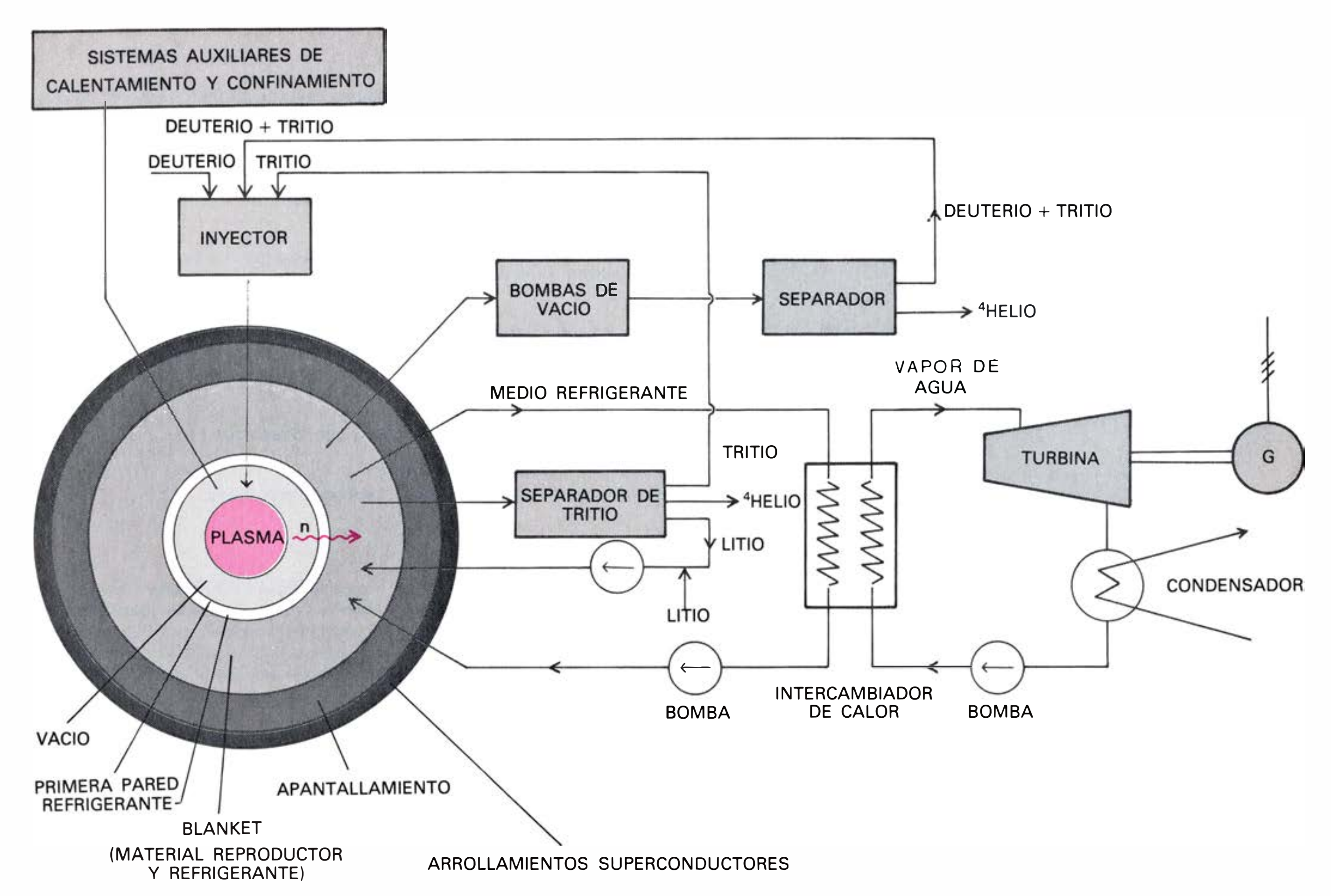
En casi todos los sistemas de confinamiento magnético estudiados hasta ahora, el plasma se encuentra confinado en un equilibrio inestable. Cualquier ligera desviación del plasma o cualquier perturbación, por pequeña que sea, en su superficie tiende a crecer, destruyendo rápidamente la configuración y precipitando el plasma sobre las paredes del recipiente que lo contiene, donde se enfría rápidamente, haciendo imposible la fusión.

Los sistemas de confinamiento magnético pueden clasificarse en sistemas abiertos y sistemas cerrados. En los sistemas abiertos, las líneas de fuerza del campo magnético salen del plasma y se cierran en el exterior del mismo; en los sistemas cerrados, esas líneas se cierran dentro del plasma o bien lo envuelven. Entre los sistemas cerrados se cuenta el tokamak [véase "Progresos hacia un reactor de fusión tokamak", por Harold P. Furth, INVESTIGACIÓN Y CIENCIA, octubre de 1979].

Y sobre el tokamak se centra hoy el principal esfuerzo de las investigaciones sobre fusión controlada, por ser el sistema que ofrece mejores perspectivas. El tokamak es un sistema de confinamiento toroidal, cerrado, desarrollado y ensayado originalmente en Rusia. En este sistema, el plasma se confina mediante dos campos magnéticos, uno toroidal, dirigido a lo largo del eje de un recipiente en forma de toro, y otro poloidal, cuyas líneas de fuerza circulares son simétricas respecto al eje del toro.

La superposición de estos dos campos da lugar a un campo magnético cuyas líneas de fuerza son helicoidales alrededor del eje toroidal y forman superficies magnéticas cerradas que envuelven el plasma. El campo está creado por corrientes externas, que circulan por arrollamientos superconductores alrededor del recipiente toroidal, y por una corriente que circula por el seno del plasma. Se induce esta última corriente por un flujo magnético creciente sobre el eje del toroide, flujo que puede originarse directamente sobre el eje; el flujo puede crearse también fuera y canalizarse luego sobre el eje por el núcleo de

CURVA CORRESPONDIENTE A LA REACCION DEUTERIO-TRITIO (D-T). Dicha curva representa el mínimo valor del parámetro de Lawson $n\tau$ (producto de la densidad del plasma, *n*, por el tiempo de confinamiento del mismo, τ), para un balance de energía positivo sobre el plasma, en función de la temperatura del mismo. Para puntos interiores a esta curva, la fusión puede autosostenerse, mientras que, para puntos exteriores, el plasma se enfría y la reacción se extingue. La mínima temperatura de ignición del plasma es de unos 5 a 6 kiloelectronvolt (1 kiloelectronvolt equivale a unos 11,6 millones de grados) y la óptima temperatura de trabajo es del orden de los 20 a 25 kiloelectronvolt; en efecto, a esta temperatura se precisa la mínima densidad y tiempo de confinamiento para mantener la ignición.

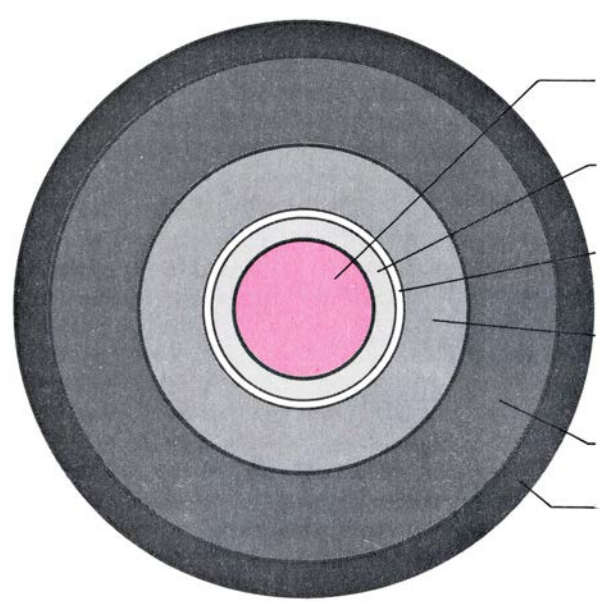


PARTES COMPONENTES de un reactor de fusión. En el núcleo, se encuentra el plasma, confinado por campos magnéticos, aislado de la primera pared por un vacío. El núcleo del reactor está rodeado de un recubrimiento o capa fértil ("blanket") de litio o de compuestos de litio, donde se regenera el tritio consumido, extrayéndose luego del material reproductor en una planta separadora, para volverlo a inyectar en el reactor. Alrededor de la capa fértil existe un apantallamiento cuya misión es atenuar las radiaciones emitidas por el plasma, y aislar térmicamente los arrollamientos superconductores que crean el campo magnético toroidal. El calor cedido por los neutrones en esa capa fértil se extrae por un medio refrigerante, que puede ser el mismo material reproductor, o bien potasio líquido, helio u otras sustancias. El medio refrigerante cede este calor en un intercambiador, generando vapor de agua que mueve una turbina, transformando su energía térmica en energía mecánica. La turbina acciona un generador de corriente alterna, produciendo corriente eléctrica. Parte de esta energía eléctrica se devuelve al reactor, donde la consumen los sistemas auxiliares de calentamiento y de confinamiento del plasma, bombas, compresores, etcétera. Unas bombas de vacío extraen continuamente, o al final de cada pulso, las impurezas, partículas de plasma difundidas y el helio 4 generado en la fusión. En una planta separadora se extraen las impurezas y el helio; el gas de D y T se vuelve a inyectar.

hierro de un transformador, en cuyo caso el anillo de plasma desempeña el papel del circuito secundario del transformador.

El gran avance de las investigaciones sobre fusión controlada con máquinas tokamak registrado en los últimos años apoya la esperanza en que pronto puedan construirse y ponerse en funcionamiento grandes reactores de fusión. Hasta el presente se ha venido ensayando con máquinas pequeñas en la investigación de los problemas físicos relacionados con la fusión; entre éstos: confinamiento de plasmas a densidades y durante intervalos de tiempo suficientes para satisfacer la condición de Lawson, calentamiento del plasma hasta la temperatura de ignición y otros. Están ya en funcionamiento aparatos experimentales algo mayores, con los que se pretende abordar problemas más complejos, como pueden ser la extracción de impurezas y partículas de plasma difundidas mediante divertores (modelos PDX en los Estados Unidos, DITE en Inglaterra y ASDEX en la República Federal Alemana), el calentamiento del plasma por inyección de haz neutro (modelo PLT en los Estados Unidos), la interacción del plasma con la primera pared del tokamak y protección adecuada de la misma (modelo TEXTOR en la República Federal Alemana), arrollamientos superconductores (modelo T-7 en la Unión Soviética), plasmas de sección no circular (modelos DOUBLET-III en los Estados Unidos, BELTPINCH en la Alemania Federal), etcétera.

Se ha llegado a la resolución de muchos problemas. Se han alcanzado, por ejemplo, temperaturas de fusión (en 1962) y se ha conseguido el confinamiento estable de plasma a densidades y durante tiempos de confinamiento suficientes para satisfacer la condición de Lawson (en 1969). Pero no se ha podido ligar esos dos logros todavía. Se han desarrollado fórmulas empíricas y teóricas para predecir la evolución de los parámetros del plasma a medida que aumenta el volumen del mismo y la intensidad del campo magnético. Nos hallamos ahora en el punto crítico de comprobar, en experimentos de fusión mayores, si esas leyes empíricas y teóricas se cumplen con los parámetros del plasma. Esa es la razón de que se estén construyendo grandes tokamaks en Europa, Estados Unidos, Rusia y Japón, cuyo funcionamiento se proyecta para los primeros años de la década que acaba de empezar. El más avanzado será probablemente el T-20 de Rusia, que consumirá como combustible una mezcla de deuterio-tritio, y dispondrá de un recubrimiento o blanket reproductor. Se espera que entre en funcionamiento en 1982. En Europa se ha iniciado la construcción



MUCHOS PROBLEMAS quedan todavía por resolver en la construcción de un reactor de fusión. Aqui se señalan los más importantes. Junto a los relativos de calentamiento y confinamiento estable del plasma, ocupan un primer lugar los concernientes a los materiales para el reactor de fusión y los problemas tecnológicos. Los problemas de materiales son esencialmente graves para la primera pared y para la capa fértil, debido a las grandes tensiones térmicas y mecánicas que sufren y al continuo bombardeo de neutrones y a las radiaciones electromagnéticas emitidas por el plasma a que se ven sometidos. Dichos ataques deterioran estas partes del reactor, por lo que habrán de sustituirse varias veces durante el tiempo de vida del mismo.

del JET (Joint European Torus), que se está instalando en Culham, Inglaterra, de apertura inmediata. En este proyecto colaboran Alemania, Francia, Inglaterra e Italia, principalmente. En Estados Unidos se está construyendo el TFTR (Tokamak Fusion Test Reactor), y en Japón el JT-60, que entrarán ambos muy pronto en funcionamiento. En todos estos tokamaks se espera poder alcanzar simultáneamente las condiciones de temperatura, densidad y tiempo de confinamiento necesarias para la fusión controlada.

El plasma pierde continuamente energía por radiación y por difusión de partículas hacia el exterior. Las pérdidas por radiación revisten importancia especial cuando el plasma está impurificado, es decir, cuando contiene iones de elevado número atómico. Estas pérdidas se producen principalmente por radiación de frenado (*Bremsstrahlung*) y por radiación de ciclotrón, así como por radiación de recombinación y de excitación. (La radiación de frenado es una radiación gamma –de fotones– que se origina en la deceleración de los electrones en el campo eléctrico de los iones. La radiación de ciclotrón es la emitida por un electrón en un campo magnético, emisión provocada por el movimiento helicoidal de esta partícula en el campo magnético.) Las pérdidas por radiación de recombinación y de excitación son prácticamente nulas a las temperaturas del plasma que consideramos. Generalmente puede despreciarse también la radiación de ciclotrón frente a la de frenado, porque es mucho más débil y además es reabsorbida en gran parte por el plasma. Sólo a temperaturas muy elevadas, como las necesarias para la fusión deuterio-deuterio, y con campos magnéticos muy intensos, esta radiación puede llegar a ser importante, superando incluso a la de frenado.

Los principales problemas en el reactor de fusión se presentan en el núcleo del reactor que contiene el plasma, rodeado de un vacío y de la primera pared. El plasma en los reactores de fusión debe ser muy puro. Las impurezas, por pequeñas que sean, aumentan grandemente las pérdidas de energía por radiación; el plasma se enfría entonces cada vez más, hasta que se extingue la reacción de fusión. Las impurezas afectan, por tanto, al tiempo durante el cual puede mantenerse la fusión en el plasma, haciendo que disminuya la potencia media del reactor. Para asegurar la pureza del plasma debe hacerse un vacío muy perfecto en el núcleo del reactor antes de inyectar el gas deuterio-tritio, y crear el plasma. Cuanto mayor es el número atómico de las impurezas, menor es la cantidad de las mismas tolerable en el plasma. Impurezas como el nitrógeno, oxígeno o carbono, de número atómico relativamente bajo, pueden estar presentes en el plasma hasta en un 1 por ciento. Sin embargo, impurezas pesadas, como el hierro, níquel y molibdeno (desprendidas de la primera pared, debido al bombardeo de neutrones y partículas de plasma difundidas) no han de superar el 0,1 por ciento, en razón de las grandes pérdidas por radiación de recombinación y radiación de excitación a que dan lugar. Un reactor de funcionamiento cuasi-estacionario tiene que disponer de medios para evitar la impurificación del plasma, así como para la extracción del helio 4 resultante de las reacciones de fusión. Ha de evitarse que las partículas energéticas del plasma, que se difunden hacia el exterior del confinamiento, alcancen la pared, donde pueden vaporizar átomos de la misma, que luego penetran en el plasma como impurificaciones.

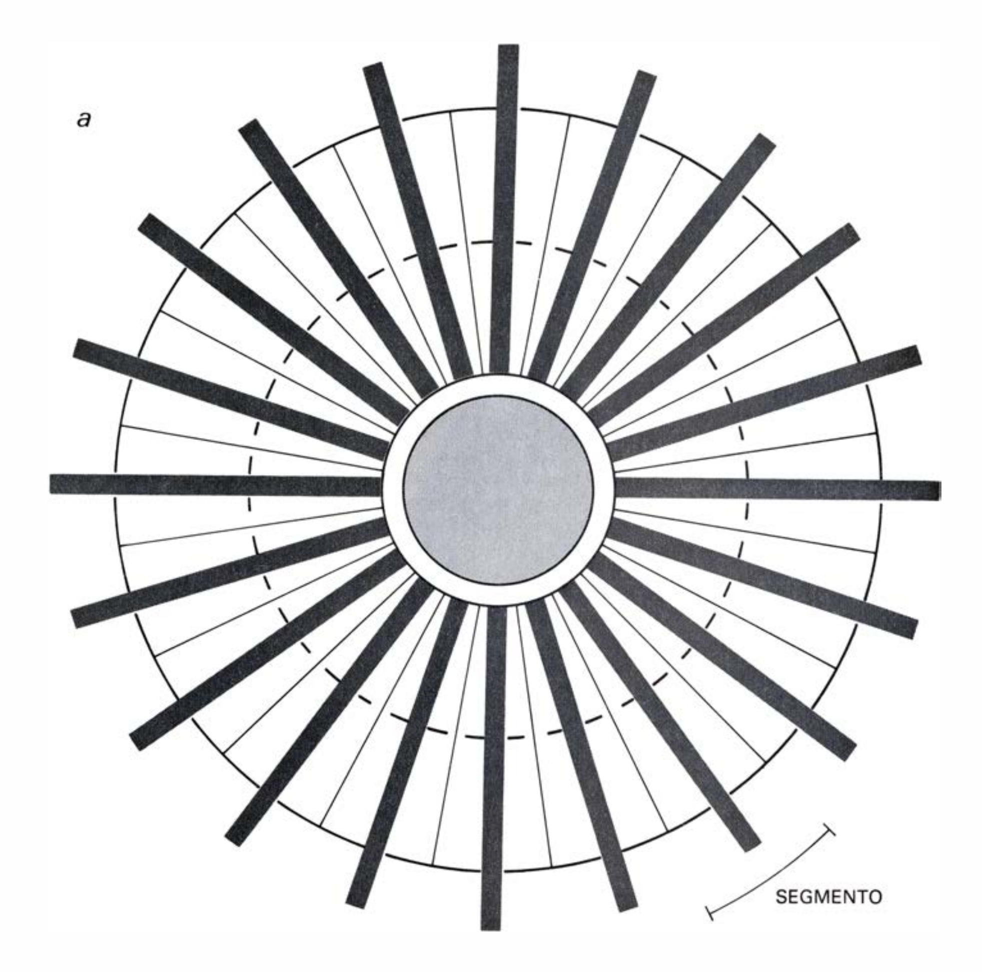
Las impurezas y partículas de plasma difundidas pueden extraerse por medio de los divertores ("divertors"). Este sistema consiste en desviar el campo magnético cerca de la superficie del plasma hacia el exterior, a través de unas aberturas en la capa fértil ("blanket") que rodea el núcleo del reactor. Para ello, en el tokamak hay unos arrollamientos, generalmente superconductores, dispuestos adecuadamente en el exterior del núcleo del reactor. Las partículas de plasma difundidas, así como las impurezas (que suelen acumularse cerca de la superficie del plasma), se canalizan entonces a lo largo de estas líneas de fuerza hacia las cámaras de bombeo situadas en el exterior del núcleo del reactor, donde son recogidas y enfriadas sobre unas placas colectoras para salir expulsadas al exterior por las bombas de vacío.

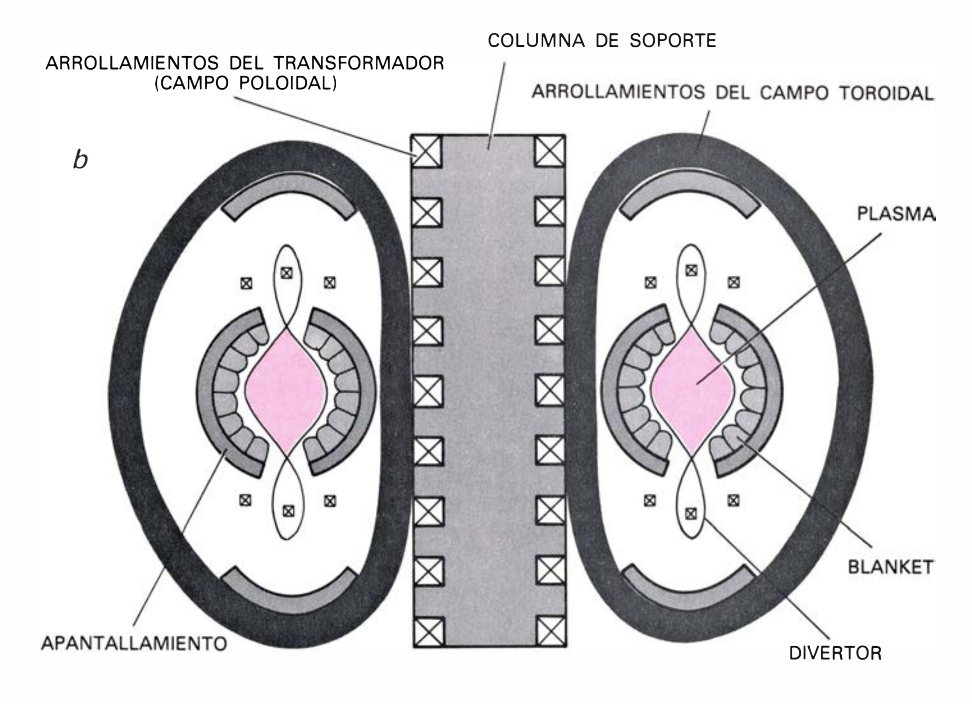
Se han ensayado varios métodos de purificación de los gases extraídos del reactor por las bombas de vacío. Se trata de la separación del deuterio, tritio y helio, así como de otras impure-

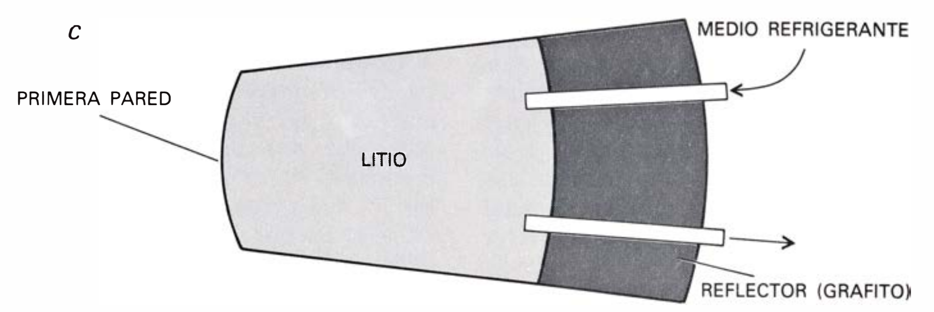
zas, argón u otros gases de número atómico elevado por ejemplo, que se inyectan al final de cada pulso para enfriar el plasma. El método de difusión a través de membranas de paladio se basa en el hecho de que el hidrógeno es capaz de atravesar paredes metálicas finas por permeabilidad, cosa que no pueden hacer el helio y otros elementos. Por procedimientos químicos deben separarse luego el deuterio y el tritio, y eventualmente también el hidrógeno que pudiera estar presente, debido a las reacciones deuterio-deuterio. En un segundo método, el de destilación criogénica, se enfría el hidrógeno, condensándose de 25 a 17 grados Kelvin, mientras el helio se separa como vapor no condensado. El hidrógeno, deuterio y tritio se separan en una pequeña columna de destilación, de unos tres centímetros de diámetro y de uno a dos metros de altura. Otros métodos se basan en la absorción del deuterio y tritio en metales en los que el hidrógeno tiene alta solubilidad, placas de circonio, aluminio y litio, por ejemplo, y separación por métodos químicos.

El alto vacío ($\simeq 10^{-5}$ torr) que debe hacerse en el núcleo del reactor antes de la inyección del gas deuterio-tritio y la creación del plasma presenta también grandes problemas: ha de evacuarse un gran volumen, del orden de varios miles de metros cúbicos. La pared de vacío puede ser la primera pared que rodea al plasma, aunque esto es poco recomendable, porque está sometida a grandes tensiones; por otro lado, debe tener un espesor pequeño, para no absorber demasiados neutrones y disminuir la razón de reproducción de tritio en la capa fértil. Otra posibilidad es hacer el vacío englobando todo ese recubrimiento o incluso todo el reactor. Con esto también se evitarían las peligrosas reacciones del litio, reproductor y refrigerante, con el aire y el vapor de agua de la atmósfera en caso de rotura de alguna conducción de litio.

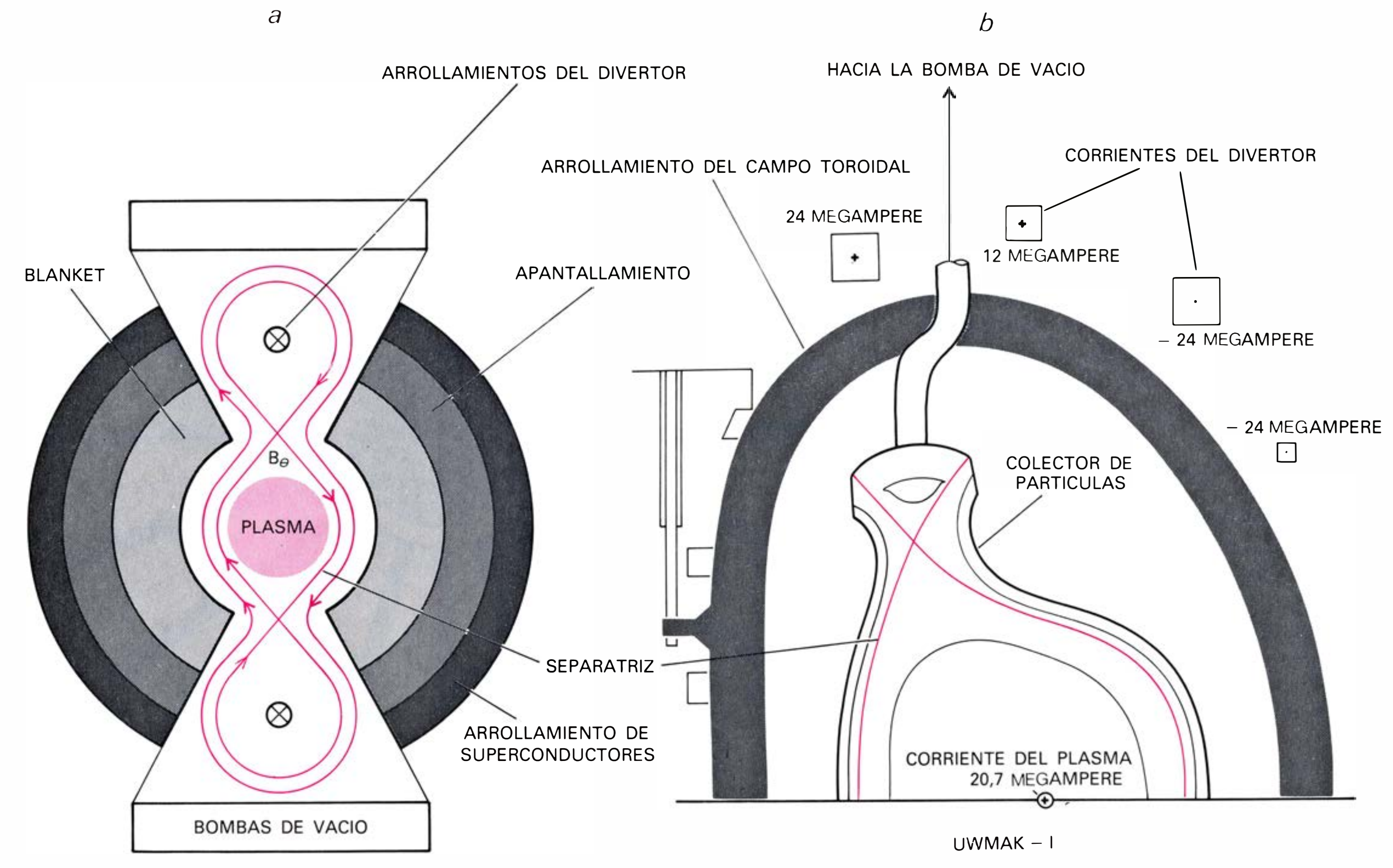
La inyección de combustible deuterio-tritio en el reactor de fusión constituye otro de los principales pro-







PARA FACILITAR LA RENOVACION o reparaciones de partes dañadas y el mantenimiento del reactor, la construcción del mismo será en forma de segmentos intercambiables, tal como se muestra en la figura (*a*), en la que se representa un reactor de fusión toroidal de veinte segmentos. Cada segmento contiene su parte correspondiente de capa fértil, apantallamiento y arrollamientos superconductores (*b*). Todas las células de la capa fértil constan de la primera pared, zona de material reproductor (litio) y refrigerante, así como de reflector y de moderador de grafito (*c*).



IMPUREZAS Y PARTICULAS DE PLASMA DIFUNDIDAS se extraen por medio de los llamados divertores. En *a* se muestra de forma muy esquemática el funcionamiento de uno de ellos. Existen dos tipos de divertores. En los poloidales, o de simetría axial, se desvía hacia el exterior el campo magnético poloidal por medio de unos arrollamientos superconductores externos, que tienen simetría axial con el toro, es decir, son paralelos a la corriente que circula por el plasma. El otro tipo de divertor es el "bundle divertor" (divertor lío), en el que se desvía hacia el exterior el campo magnético

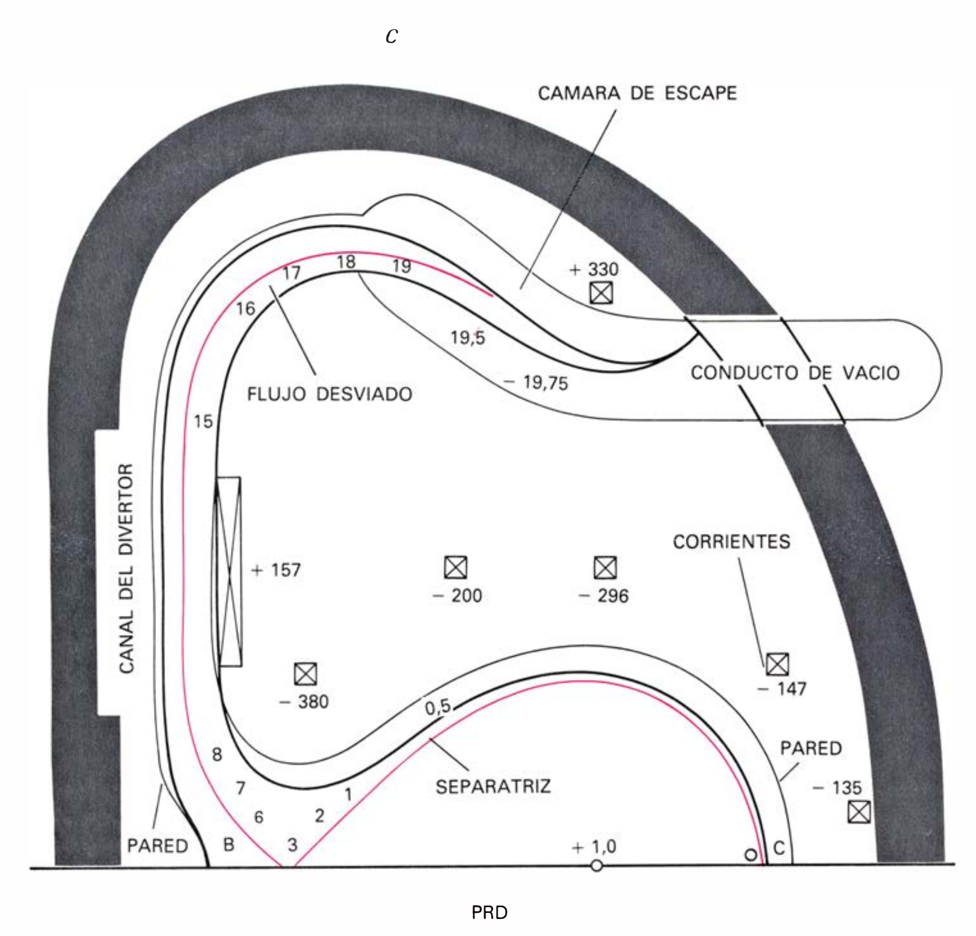
blemas que quedan por resolver para la construcción de un reactor de fusión. El combustible debe inyectarse en estado neutro, pues las partículas cargadas no pueden atravesar el campo magnético que rodea al plasma. Los átomos de deuterio y tritio han de inyectarse con la energía necesaria para que puedan penetrar con suficiente profundidad en el plasma, antes de ionizarse por choques con las partículas de éste, y quedar atrapados por el campo magnético.

Se han propuesto varios métodos de inyección de combustible. Uno de ellos es la inyeción de haz neutro. Gracias a la gran energía de las partículas neutras inyectadas ($\simeq$ 500 kiloelectronvolt), su recorrido libre medio en el plasma es grande, por lo que pueden penetrar en éste profundamente antes de quedar ionizadas y atrapadas en el campo magnético. Incluso puede regularse el lugar de deposición de las partículas inyectándolas con mayor o menor energía. Sin embargo, el método de inyección de haz neutro no resulta muy aconsejable porque consume demasiada energía y, debido a esa gran energía con que deben inyectarse las partículas de deuterio y tritio para que puedan llegar al centro del plasma, éste se sobrecalentaría, lo que redundaría en una disminución de la potencia de fusión. Sólo en el reactor de espejos magnéticos, que trabajaría con plasmas a gran temperatura, podría utilizarse la inyección de haz neutro para el calentamiento del plasma y para la inyección de combustible.

Más prometedor parece ser el método de inyección de combustible en forma de gotas de deuterio-tritio sólidas. Estas gotas deben inyectarse a gran velocidad, para que lleguen al centro del plasma antes de su total evaporación. Cuando la gota ("pellet") penetra en el plasma, los átomos de su superficie se ionizan, formando una capa de plasma muy densa, de baja temperatura, alrededor de la gota, que la protege, lo que permite que alcance el centro del plasma antes de haberse evaporado. El diámetro de las gotas puede variar desde la micra hasta el milímetro; las velocidades de inyección requeridas son del orden de 10.000 metros por segundo, según la temperatura y densidad del plasma y la profundidad de penetración que se desee. Los sistemas de aceleración de estas gotas están, sin embargo, poco estudiados.

También se ha propuesto el aporte de combustible al plasma por un recubrimiento de gas. Hay alrededor del plasma una capa de gas deuterio-tritio frío, de densidad relativamente grande. Esta capa de gas, además de proteger la primera pared contra el impacto de las partículas difundidas, suministraría nuevo combustible al plasma, por difusión de los átomos de deuterio-tritio hacia el interior del plasma. El helio 4 resultante de la fusión, cuya densidad es mayor en el centro del plasma, se difundiría hacia el exterior. Con este método existirían sin embargo grandes pérdidas por transporte de energía por conducción.

Los materiales de estructura del reactor de fusión están sometidos a intensas radiaciones, que provocan el deterioro de sus propiedades físicas, grandes tensiones térmicas y una fuerte radiactividad inducida en los mismos. Estos problemas se agravan en la primera pared, debido a la interacción directa de la misma con el plasma, y al mayor flujo neutrónico que la atraviesa. La primera pared está sometida a radiaciones electromagnéticas emitidas por el plasma (radiación de frenado,



toroidal. Este divertor no tiene simetría axial como el anterior. Consiste en esencia en un arrollamiento solenoidal colocado perpendicularmente a la superficie del plasma en algún punto de la misma. En *b* y *c* aparecen los divertores de dos importantes diseños conceptuales: el diseño UWMAK-I, de la Universidad de Wisconsin, y el diseño PRD, de la Universidad de Princeton. Ambos son de tipo poloidal.

radiación de ciclotrón, radiación de recombinación y radiación de excitación) y a radiación gamma inducida por los neutrones en ella y en otros materiales de la estructura del reactor. Esa primera pared sufre también el bombardeo de los neutrones de 14 megaelectronvolt resultantes de las distintas reacciones de fusión (unos $10^{14} - 10^{15}$ neutrones/centímetro cuadrado × segundo), y es blanco del bombardeo de las partículas de plasma difundidas (electrones, iones y partículas alfa) y de átomos neutros energéticos. Estos átomos neutros energéticos resultan de los iones del plasma difundidos, que se neutralizan por intercambio de carga, al chocar con átomos de gas neutro en el exterior del plasma.

Las radiaciones electromagnéticas emitidas por el plasma son absorbidas en gran parte en la primera pared, que se calienta fuertemente (500 a 1000 grados Celsius), por lo que debe refrigerarse adecuadamente para evitar que su temperatura se eleve peligrosamente. Ese salto de temperatura entre la superficie interior y la exterior de la primera pared determina la aparición en ella de grandes tensiones térmicas. Por otro lado, su espesor debe ser pequeño (de 0,4 a 1 centímetros) para evitar que se absorban demasiados neutrones, y que se reduzca con ello la razón de reproducción de tritio en el recubrimiento. La relevancia de las tensiones térmicas en la primera pared se aprecia en seguida si tenemos en cuenta que la atraviesa una potencia de 1 a 4 megawatt por metro cuadrado, engendrada por las radiaciones y los neutrones emitidos por el plasma.

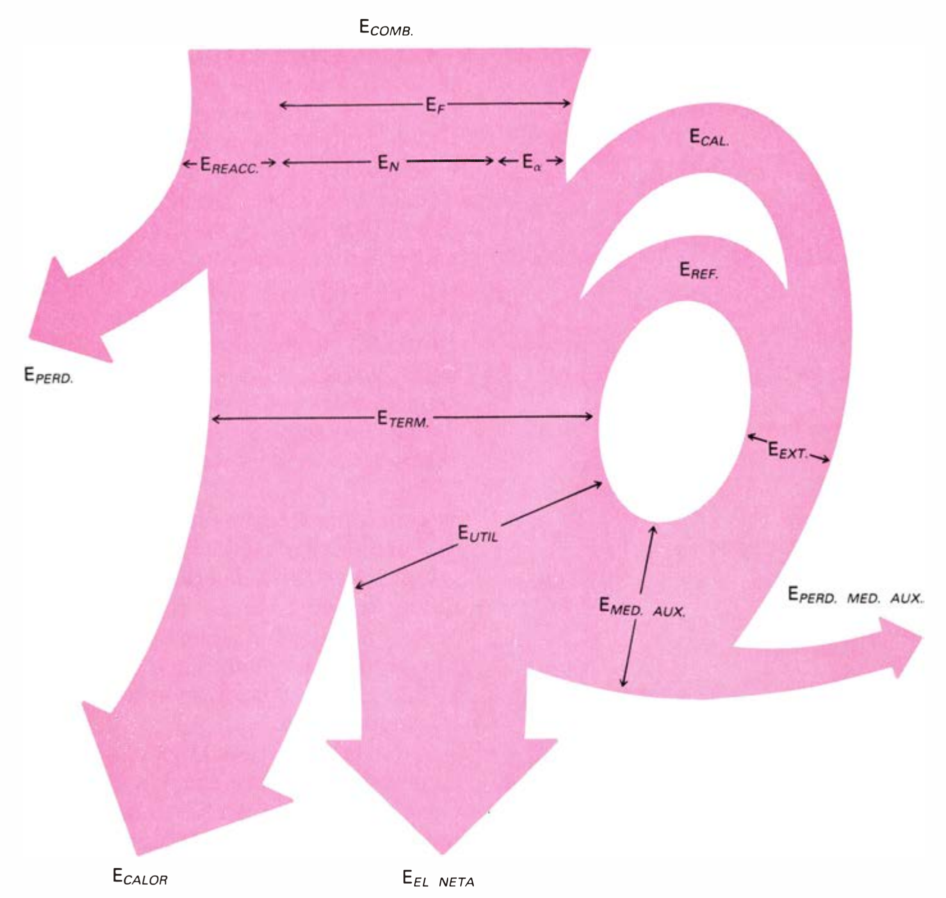
Al chocar con los átomos de la primera pared y demás materiales de estructura del reactor, los neutrones rápidos dan lugar a desplazamientos atómicos, en virtud de los cuales se producen oquedades en la red cristalina que debilitan la resistencia del material y se alteran sus propiedades físicas, causan también transmutaciones de unas partículas en otras y determinadas reacciones. Las principales consecuencias del bombardeo neutrónico en la primera pared pueden resumirse en las siguientes: hinchamiento de la pared, pérdida de ductibilidad que se manifiesta a través de resquebrajaduras y fracturas, disminución de su resistencia a los esfuerzos mecánicos, radiactividad inducida y erosión de ese muro en cuestión.

El hinchamiento de la primera pared y demás materiales de estructura del reactor se debe a la formación de bolsas de helio por reacciones del tipo neutrones-partículas alfa, en las que se forman hidrógeno y helio. En tanto que el hidrógeno es muy soluble en la mayoría de los metales, sobre todo a elevada temperatura, y puede difundirse hacia el exterior, el helio queda atrapado, formando pequeñas burbujas y huecos en el material. La magnitud del hinchamiento de los materiales por bombardeo neutrónico depende de la temperatura de aquéllos.

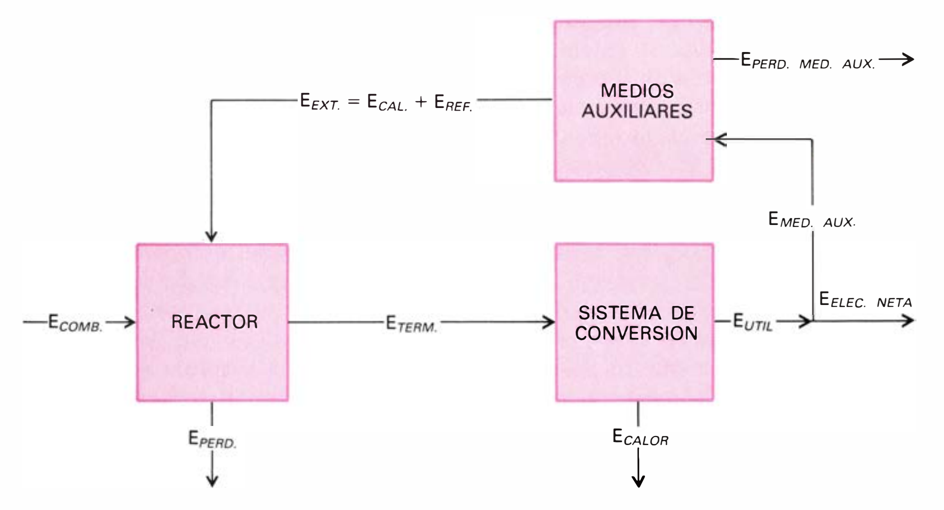
Dos consecuencias de ese bombardeo, la pérdida de ductibilidad y la disminución de resistencia de los materiales, resultan de la formación de huecos en la red cristalina por desplazamientos de los átomos que la forman, al chocar con ellos los neutrones, y de transmutaciones.

Los neutrones que chocan con los átomos de la primera pared pueden dar lugar a reacciones nucleares que generen nuevos núcleos radiactivos. Se trata del fenómeno conocido por radiactividad inducida, que afecta asimismo a los materiales de estructura de recubrimiento, sometidos a un intenso flujo de neutrones y, en menor medida, al apantallamiento y a los materiales reproductores y refrigerantes. La radiactividad inducida puede ser peligrosa en caso de un accidente en el sistema de refrigeración, debido al calor residual que desarrollan los materiales activados, que puede ser hasta de un 0,2 a un 2 por ciento de la potencia térmica del reactor.

La erosión limita el tiempo de vida de la primera pared, que deberá cambiarse cada uno o dos años, y ocasiona una impurificación del plasma. En la erosión de la primera pared intervienen, junto a los neutrones, las partículas de plasma que se difunden hacia la pared y los átomos neutros energéticos. Puede producirse de varias formas: por evaporación, por pulverización ("sputtering") y por pustulación ("blistering"). La evaporación de átomos de la primera pared se produce cuando la temperatura de ésta aumenta de suerte que sus átomos adquieren suficiente energía como para vencer la fuerza de ligadura que los mantiene unidos a la superficie; y así se desprenden de ésta. La pulverización se produce cuando un neutrón, ion, o átomo neutro energético choca contra un átomo de la primera pared. Este átomo recibe entonces energía suficiente para desprenderse de la pared; puede además interaccionar fuertemente con



BALANCE GLOBAL DE ENERGIA sobre el reactor de fusión, resumido en un diagrama de flujo de energía. En el dibujo, e_{comb} es la energía liberada por los combustibles consumidos en el reactor (deuterio y litio, en definitiva). Las reacciones deuterio-tritio que tienen lugar en el plasma liberan una energía e_f, de la cual un 80 por ciento es energía de los neutrones, y un 20 por ciento energía de los núcleos de helio 4 o partículas alfa. Mientras que las partículas alfa permanecen en el plasma, los neutrones, sin carga, salen del confinamiento y pasan a la capa fértil donde reaccionan con el litio y en virtud de lo cual se libera, en promedio, cierta energía e_{reacc}. En el diagrama de flujo e_{cal} es la energía invertida desde el exterior para el calentamiento del plasma (calentamiento óhmico, inyección de haz neutro, etcétera); e_{ref} representa la energía térmica del medio refrigerante a la entrada del reactor; e_{term} simboliza la energía térmica del medio refrigerante a la salida del reactor; e_{perd} designa la energía que se pierde a través del apantallamiento (por conducción de calor, radiación no absorbida, y otros). La energía térmica del medio refrigerante e_{term}, se convierte en otras formas de energía útil, por un sistema de conversión. Podemos convertir parte de la energía térmica del refrigerante en energía útil, el resto se disipa como calor.



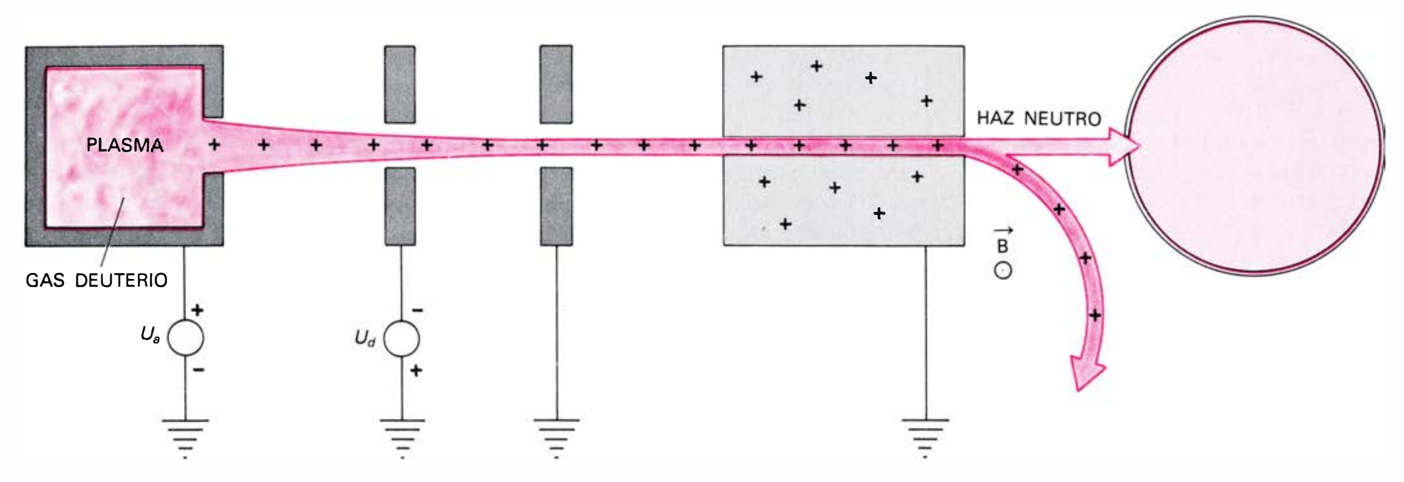
ESTE ESQUEMA detalla el esquema de flujo anterior. De la energía útil obtenida, parte debe utilizarse para mantener el funcionamiento del reactor a través de los medios auxiliares $e_{med.\ aux.}$ (sistema de calentamiento y confinamiento del plasma, refrigeración de arrollamientos superconductores, sistemas de bombeo, y demás), el resto es energía eléctrica neta, $e_{el.\ neta}$. De la energía $e_{med.\ aux.}$, parte torna al reactor como energía de calentamiento del plasma (e_{cal}) y como energía del medio refrigerante (e_{ref}); el resto se disipa en forma de calor en los sistemas de calentamiento, de bombeo, etcétera, y lo englobaremos en $e_{perd.\ med.\ aux.}$. En definitiva, de la energía e_{comb} que se genera en el reactor, sólo una parte se obtiene como energía eléctrica neta, $e_{el.\ neta}$, y el resto se disipa por pérdidas a través del apantallamiento, e_{perd}, a través de los medios auxiliares, $e_{perd.\ med.\ aux.}$, o calor residual en conversión e_{calor}. El reactor de fusión puede considerarse como un "amplificador de energía". Se invierte en él cierta energía e_{ext}, y se obtiene a cambio una energía $e_{ext} + e_{comb}$, siendo e_{comb} la energía liberada por los combustibles en el reactor.

otros átomos, arrastrándolos consigo. La pustulación resulta de la formación de ampollas de gas en la superficie de la primera pared. Este fenómeno se debe principalmente al bombardeo de partículas de plasma que sufre la pared. Las partículas penetran dentro de la primera pared algunos angstrom, gracias a su gran energía. El deuterio y tritio pueden difundirse a través del metal y salir otra vez a la superficie; no así el helio, que queda atrapado en la primera pared, formando pústulas ("blisters") en la superficie de la misma. Después de cierto tiempo, esas ampollas revientan, originando una fuerte erosión de la primera pared.

El deterioro de los materiales causado por bombardeo de neutrones está todavía poco estudiado experimentalmente, ya que es difícil disponer de fuentes de neutrones a 14 megaelectronvolt.

La primera pared puede protegerse contra las partículas de plasma difundidas mediante un divertor. Sin la existencia de un divertor, que desvíe las partículas de plasma difundidas hacia las cámaras de bombeo antes de que alcancen la primera pared, el flujo de partículas que llegaría a ésta sería de unos 10^{23} iones por segundo para un reactor de 5000 megawatt de potencia térmica. Ello provocaría una erosión de la primera pared de aproximadamente un milímetro por año. Incluso con divertor, los neutrones y átomos neutros producen una fuerte erosión por pulverización, del orden de varias décimas de milímetro por año. Se han ensayado diversos métodos para proteger el muro. Consiste uno en recubrir esa primera pared con una capa de material de bajo número atómico, carbono por ejemplo. La impurificación del plasma por carbono es menos grave que la de niobio, molibdeno, hierro o níquel, materiales que usualmente se eligen para la primera pared. Según otro método, el muro se salvaguarda mediante un recubrimiento de gas, una capa de gas deuterio-tritio frío, de densidad relativamente elevada, que, al tiempo que sirve de protector, se emplearía como medio de aporte de combustible.

La elección del material adecuado para la primera pared constituye uno de los puntos decisivos del reactor de fusión. En los modelos ideales diseñados hasta el presente se han considerado materiales como el acero, el níquel, el niobio, el molibdeno, el vanadio y el aluminio. El material ideal debe satisfacer las siguientes condiciones: no sufrir un deterioro demasiado rápido



LA INYECCION DE HAZ NEUTRO es un sistema de calentamiento para el tokamak. Un inyector de haz neutro funciona así: se acelera electrostáticamente un haz de iones, que se extraen de una fuente de plasma, y se hace pasar por una célula de gas neutro, donde en parte se neutraliza por intercambio de cargas. La parte de haz neutro no neutralizada se desvía por medio de un campo magnético y se recupera su energía por conversión directa. El resto se inyecta en el plasma, donde se ioniza por choques con las partículas del plasma quedando atrapado por los campos magnéticos.

por la radiación de neutrones, su radiactividad inducida se mantendrá en niveles bajos para decaer rápidamente, apenas contaminará el plasma, permitirá trabajar a grandes temperaturas, soportar grandes tensiones y resultar de bajo coste y fácil elaboración.

Ninguno de los materiales citados reúne todas esas características, por lo que habrá que desarrollar nuevas aleaciones y ensayar muchos proyectos experimentales. La ventaja del uso del acero reside en la gran experiencia que se posee de la fabricación y elaboración de ese material. Sin embargo, su techo de temperatura de trabajo se halla entre los 500 y los 700 grados Celsius. El niobio y el molibdeno tienen en su favor que pueden trabajar a grandes temperaturas (hasta más de 900 grados Celsius). En muchos diseños, la primera pared está compuesta de niobio en aleación con circonio. Aluminio y vanadio, por su parte, presentan la ventaja de una baja radiactividad inducida, que decae rápidamente. En todo caso es de esperar que sea necesario sustituir la primera pared, así como otras partes del reactor, varias veces durante el tiempo de vida (20-30 años) del mismo.

El recubrimiento o capa fértil que rodea el núcleo del reactor de fusión tiene dos funciones: primera, moderar los neutrones de 14 megaelectronvolt resultantes de las reacciones de fusión y absorber su energía por un medio refrigerante; y, en segundo lugar, reproducir tritio consumido en la fusión, por la reacción de los neutrones con el litio, o con compuestos de litio, que contiene el recubrimiento. El número de átomos de tritio producidos por neutrón absorbido en la capa fértil, o lo que es lo mismo, la relación entre el número de átomos de tritio producidos y el número de átomos de tritio consumidos, se llama razón de reproducción. Es necesario que esta razón de reproducción sea al menos igual a la unidad para que no haya un consumo neto de tritio. Sin embargo, una razón de reproducción mayor que la unidad es interesante porque se consigue así más tritio del que se ha consumido, pudiéndose utilizar el excedente para poner en marcha nuevos reactores de fusión. La razón de reproducción puede ser mayor que la unidad gracias a la reacción endotérmica expuesta al principio, en que se unía litio 7 con neutrones para formar helio 4, tritio y neutrones; en ella, además de generarse un átomo de tritio, se libera un neutrón lento, que puede volver a reaccionar con el litio por la reacción litio 6 + neutrones. Para asegurar una razón de reproducción mayor que la unidad, la capa fértil puede contener también materiales "multiplicadores" de neutrones, como por ejemplo el berilio. La reacción del berilio 9 con un neutrón da lugar a dos núcleos de helio 4 y a dos neutrones. Como multiplicadores de neutrones se utilizan generalmente el berilio o el óxido de berilio. También presentan una gran sección eficaz a las reacciones neutrón, dos neutrones $(n, 2n)$ el plomo y el niobio.

Más que la razón de reproducción, la magnitud que realmente describe la eficacia del reactor, como reproductor de tritio, es el tiempo de doblaje. Se entiende por tal el tiempo que tarda en conseguirse una cantidad de tritio doble de la que contenía el reactor inicialmente, de forma que pueda ponerse en funcionamiento un nuevo reactor de fusión de potencia igual a la del anterior. El tiempo de doblaje depende de la potencia del reactor; cuanto mayor sea ésta, aunque se consuman más átomos de tritio, también se producen más neutrones, pudiéndose generar, por las reacciones 5 y 6 nuevos átomos de tritio. El tiempo de doblaje se halla en relación inversamente proporcional a la potencia térmica del reactor, con el mínimo inventario inicial de tritio, y a la razón de reproducción.

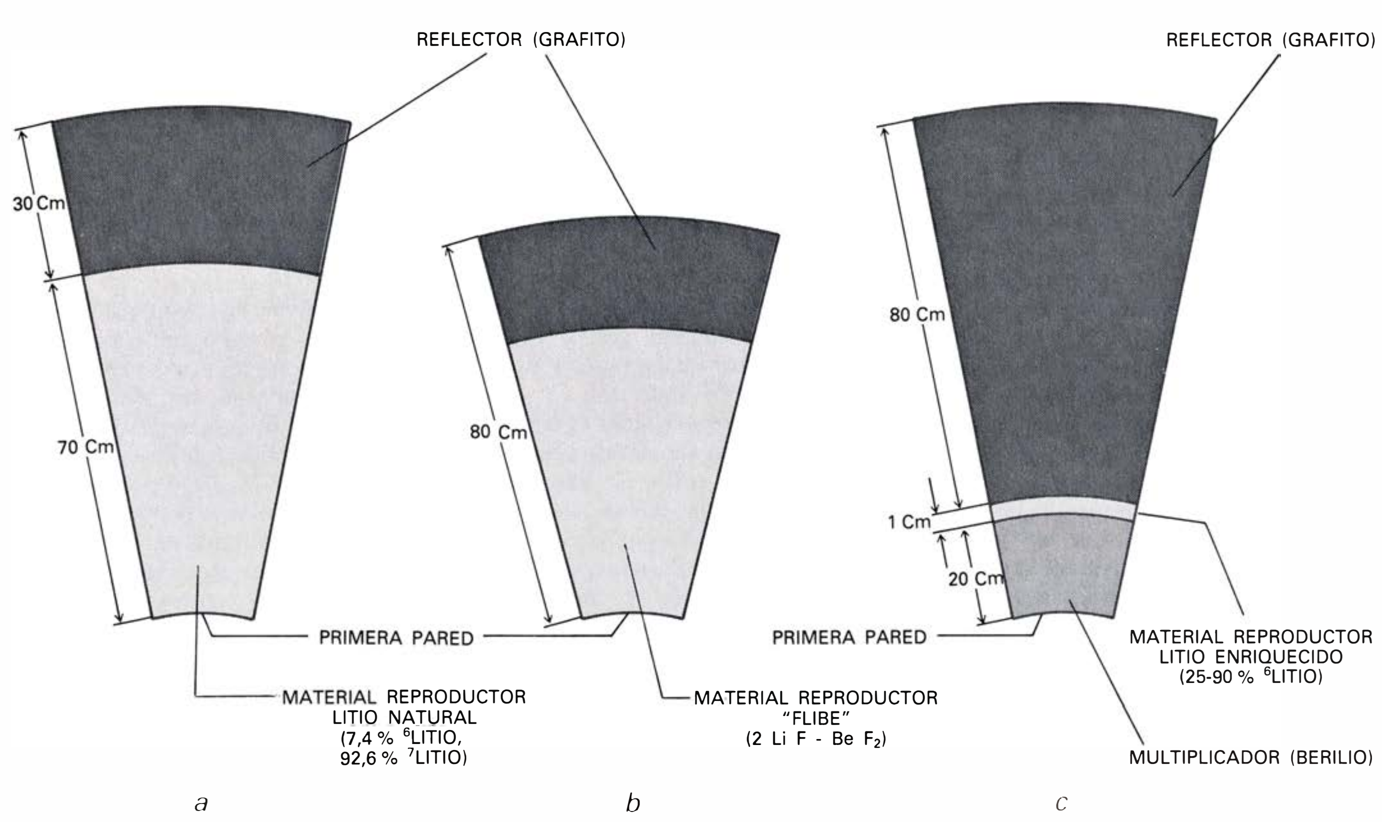
Otros motivos justifican que el inventario de tritio presente en el reactor sea lo más reducido posible. En primer lugar, su coste muy elevado; en segundo lugar, porque se trata de una sustancia radiactiva volátil, que encierra un alto riesgo de peligrosidad desde el punto de vista de la contaminación. Cuanto más tritio haya en el reactor, mayor posibilidad de contaminación por permeabilidad a través de las paredes, por fugas y escapes en caso de accidente. La cantidad de ese isótopo radiactivo que encontramos en un comienzo en el reactor es suma de las siguientes fracciones: contenido en el plasma (una cantidad mínima), tritio de la capa fértil y el sistema de recuperación del tritio, el que existe en el sistema de bombas de vacío y purificación del gas deuterio-tritio no consumido y, por último, la fracción correspondiente a los depósitos de combustible, inyectores y demás. De todos los sumandos, la porción más abundante se halla en la capa fértil o blanket y en el sistema de recuperación del tritio.

La razón de reproducción depende del espesor de la región de reproducción en la capa fértil, del material que la compone, del berilio u otros multi-

plicadores de neutrones que contenga, del material de la primera pared y materiales de estructura, del material moderador, etcétera. Desde el punto de vista de reproducción del tritio podemos distinguir tres tipos de capa fértil por razón del material reproductor empleado: los "blanket" que usan litio natural como material reproductor, las que prefieren una mezcla de fluoruro de litio y fluoruro de berilio y las que emplean litio-7 enriquecido en litio-6. En algunos diseños se ha propuesto también la utilización de compuestos sólidos derivados del litio (que suele ir enriquecido con Li-6) y el aluminio como material reproductor; en cuyo caso el material sólido se adapta a una forma esférica o bien va en barras recubiertas por una vaina metálica. La ventaja de este último diseño sería su baja concentración de tritio. No acaba ahí la lista de materiales reproductores sugeridos. Compuestos de litio con carbono, silicio, plomo y el peróxido de litio se han recomendado por su baja estabilidad térmica y escasa solubilidad del T en ellos.

La separación del tritio del material reproductor requiere una tecnología muy avanzada y compleja. Entre las propiedades del tritio está la de poder atravesar paredes metálicas por permeabilidad, sobre todo si éstas se encuentran a elevada temperatura, por lo que la densidad del isótopo debe mantenerse en niveles lo más bajo posibles en todo el reactor. Ello dificulta la separación. Hay en estudio varios métodos para la extracción del tritio. Uno de ellos consiste en aprovechar su capacidad de difundirse a través de paredes metálicas, es decir, su permeabilidad. El tritio se extrae del material reproductor (litio o una mezcla de fluoruro de litio y fluoruro de berilio) haciendo pasar a éste por unos tubos de paredes metálicas delgadas, de niobio o vanadio, a través de las cuales puede difundirse el tritio. La rapidez de difusión es directamente proporcional al área de difusión y a la permeabilidad de la pared (que depende del material de que se trate y de la temperatura) e inversamente proporcional al espesor de la pared. Depende también de las presiones parciales del tritio a ambos lados de la pared. Este sistema de extracción requiere una concentración relativamente grande de tritio en el material reproductor, porque la presión parcial del tritio en el litio es muy pequeña. Por otra parte, la concentración de tritio debe mantenerse lo más baja posible para evitar el riesgo de contaminación radiactiva, sobre todo cuando el material reproductor actúa a la vez de refrigerante, pues entonces el tritio puede pasar directamente al medio ambiente por permeabilidad a través de los intercambiadores de calor.

De acuerdo con otro método, el tritio se extrae, por enfriamiento del litio hasta temperaturas próximas al



EN UNA CAPA FERTIL DE LITIO NATURAL se tiene una razón de reproducción superior a uno, gracias a la reacción de litio-7 con neutrones. El litio actúa a la vez como material moderador, lo que es posible debido a su bajo peso atómico. El grosor de la capa fértil viene determinado por la moderación de los neutrones de 14 megaelectronvolt, y debe ser de alrededor de un metro. Consiste en una zona de aproximadamente 70 centímetros de espesor de litio natural (moderador y reproductor), seguida de otra zona de 30 centímetros de espesor de grafito (moderador y reflector). En algunos diseños también se propone utilizar acero inoxidable. Cuando la capa fértil usa como material reproductor un compuesto de fluoruro de litio y fluoruro de berilio (flibe), éste puede tener un grosor menor, unos 80 centímetros, porque absorbe más rápidamente los neutrones. La razón de reproducción es mucho menor que en la capa fértil de litio natural, debido a la baja densidad de litio en el flibe; bastan razones de reproducción, sólo algo mayores que la unidad, para tener tiempos de doblaje aceptables. La ventaja de utilizar el flibe como material reproductor del tritio estriba en que la extracción del tritio es mucho más fácil que en el litio metal, debido a la formación de fluoruro de tritio, relativamente insoluble y fácil de extraer. En la tercera capa fértil que consideramos se utiliza como material reproductor litio enriquecido en un 25-90 por ciento de litio-6, en vez de litio natural. Esta capa fértil se llama también blanket térmico, porque el tritio se genera principalmente en la reacción exotérmica de litio-6 con neutrones lentos (para dar helio-4 y tritio). Los neutrones se moderan en una gruesa capa de grafito (de unos 80 centímetros), que actúa de reflector. Para conseguir una razón de reproducción mayor que la unidad, la capa fértil contiene primero una zona de unos 20 centímetros de grosor de berilio o de óxido de berilio que actúa de multiplicador de neutrones. Debido a la gran sección eficaz de la reacción de litio-6 con neutrones lentos es suficiente un espesor del orden del centímetro de la región de reproducción. Esta capa fértil tiene la ventaja de un bajo inventario de litio, y, por tanto, también de tritio. Esto es muy favorable por razones de economía (elevado costo del tritio), por razones de seguridad (radiactividad) y porque se reduce considerablemente el tiempo de doblaje.

punto de solidificación, en forma de trituro de litio sólido. Pero este proceso requiere también una concentración mínima de tritio en el litio y, a pesar de que la presión de vapor del litio es mayor que la presión parcial del tritio, parece viable separar el tritio por destilación si atendemos a la alta solubilidad de éste en el litio y a su extensa dilución. Este método permitiría la extracción del tritio con concentraciones mucho más bajas del mismo.

La capa fértil debe enfrentarse también con el problema de su refrigeración. Se ha de eliminar el calor depositado en ella por los neutrones y por las radiaciones emitidas por el plasma. Un buen medio refrigerante tendría que reunir las siguientes características: una alta conductividad térmica y elevado calor específico, elevada densidad, bajo punto de fusión y alto de ebullición si se trata de un líquido, no ser corrosivo, no reaccionar con el aire o vapor de agua de la atmósfera y no excitarse fácilmente por los neutrones emitidos por el plasma. ¿Cómo refrigerar pues la capa fértil? Podría seguirse un procedimiento directo a través del mismo material reproductor (litio natural) o la mezcla de fluoruro de litio y fluoruro de berilio. O bien un camino indirecto mediante otros elementos: helio, potasio líquido, etcétera.

La refrigeración directa por litio líquido ofrece la ventaja de su sencillez, elevada conductividad térmica y alto calor específico, elevada densidad y bajo punto de fusión (180 grados centígrados) y alto punto de ebullición (1380 grados centígrados) del litio. Pero también presenta grandes inconvenientes. En primer lugar se trata de un elemento muy corrosivo para los materiales de estructura, especialmente para aleaciones a base de níquel. El uso del litio como refrigerante requiere temperaturas de trabajo menores que con otros medios refrigerantes para evitar una corrosión excesiva. Otro inconveniente reside en que reacciona muy fácilmente, y de forma explosiva, con el agua y el aire; lo que representa un peligro real en caso de rotura de una conducción de transporte. La principal desventaja de la refrigeración por litio quizás estribe en que, por tratarse de un metal, y por tanto ser eléctricamente conductor, interacciona con el campo magnético al moverse a través del mismo y se producen caídas de presión de origen magnetohidrodinámico.

La mezcla de fluoruro de litio y fluoruro de berilio (conocida por su contracción simbólica de "flibe") aventaja al litio en ser menos corrosiva y reaccionar menos fácilmente con el agua o con el aire. Pero tiene las contrapartidas de su elevado punto de fusión (455 grados C) y una radiactividad inducida relativamente grande.

Esas consideraciones inclinan la balanza hacia la refrigeración indirecta por helio. El helio es un gas inerte. No plantea, pues, problemas de corrosión ni de irradiación por los neutrones, ni de interacción con el campo magnético al no ser elemento conductor. Por tratarse de un gas presenta el inconveniente de su baja densidad y precisa, por tanto, de una mayor potencia de bombeo. Además, en la refrigeración por helio los materiales de la capa fértil estarán sometidos a temperaturas mayores que en la refrigeración por litio o flibe. El helio atraviesa el blanket por unos conductos, bañados exteriormente por el material reproductor (litio), que están distribuidos de suerte que cada conducto reciba igual cantidad de calor. (Los conductos de refrigeración están más próximos entre sí en la parte interna de la capa fértil, que recibe mayor cantidad de energía, que en la parte externa, más alejada de la primera pared.) Los conductos están orientados paralelamente al plasma, bien en dirección toroidal, o bien en dirección poloidal. Cada célula de la capa fértil se refrigera independientemente, aunque todas se hallan conectadas a un conducto de entrada de helio frío y a un conducto de salida de helio caliente. La conversión de energía puede hacerse por una turbina de helio (ciclo de Brayton), si el helio sale del reactor a temperatura suficientemente elevada ($\simeq$850 grados C), o generando primero vapor de agua en un intercambiador y por una turbina de vapor (ciclo Rankine). Por último, el potasio como medio refrigerante presenta la ventaja de su elevada densidad y de su bajo punto de fusión (63 grados C).

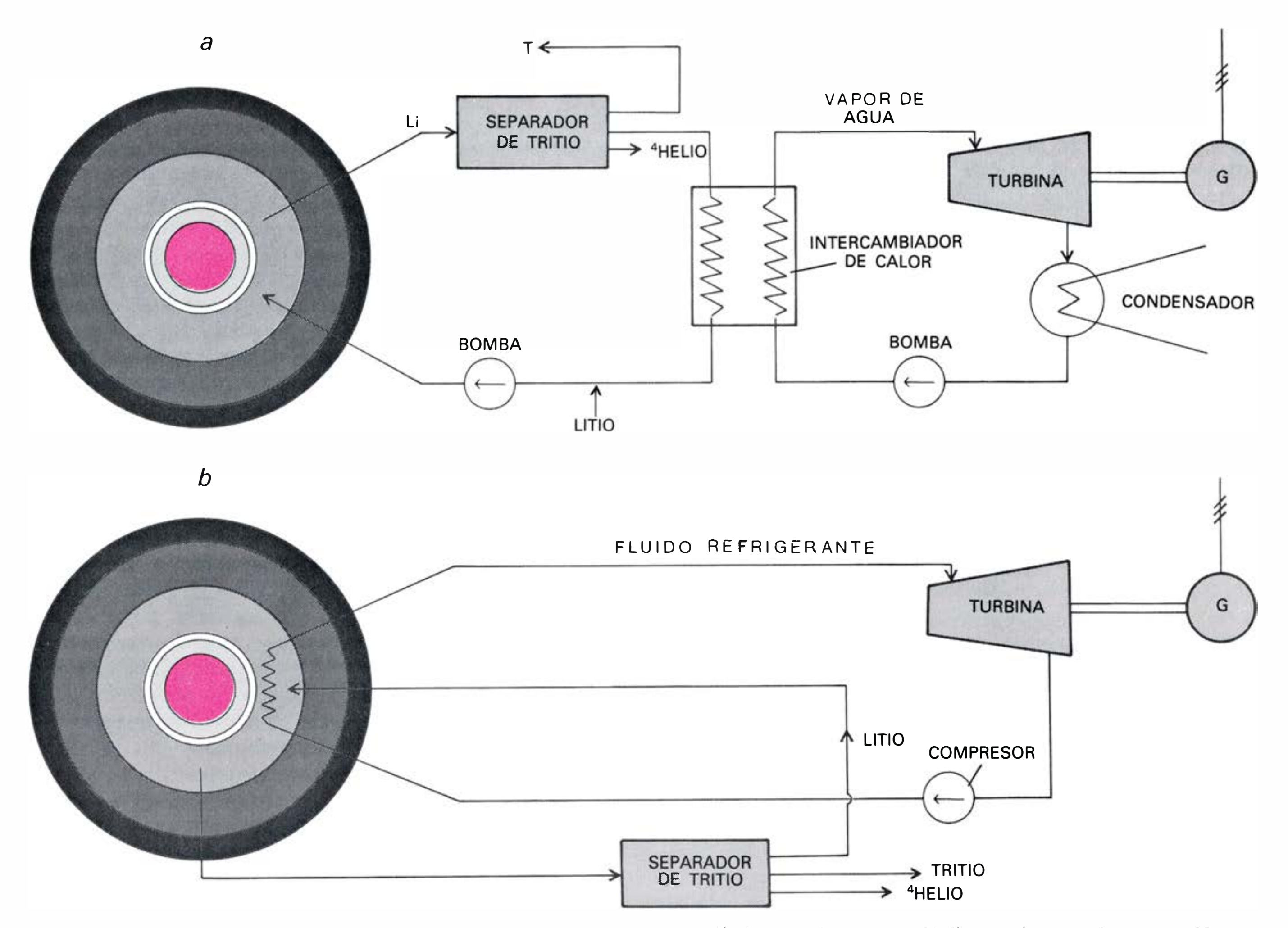
El apantallamiento que rodea a la capa fértil sirve para atenuar las radiaciones emitidas por el plasma y aislar térmicamente los arrollamientos superconductores. Con un grosor de 50 a 100 centímetros, suele ser de acero inoxidable, plomo y carburo de boro (para la absorción de las radiaciones electromagnéticas y de neutrones); se refrigera por agua o por helio. Para asegurar el aislamiento térmico de los arrollamientos superconductores, entre éstos y el apantallamiento media un espacio de aproximadamente 1 cm de ancho, en el que se ha hecho el vacío. También se utilizan mylar y materiales orgánicos para el aislamiento térmico.

Los arrollamientos que crean el campo magnético toroidal en el reactor de fusión deberán ser superconductores, porque por ellos han de circular corrientes de alrededor de 10 kiloampère para crear campos magnéticos de más de 10^4 gauss en el núcleo del reactor. Si se utilizasen arrollamientos de cobre refrigerados por agua, por ejemplo, consumirían una gran parte de la potencia eléctrica obtenida del reactor, debido a las pérdidas por efecto Joule que se producirían en ellos. Parece, pues, más rentable utilizar arrollamientos superconductores.

Los arrollamientos superconductores están constituidos por materiales que tienen la propiedad de que, cuando se enfrían hasta temperaturas próximas al cero absoluto, su resistividad decrece bruscamente hasta cero, o al menos se hace tan pequeña que puede circular por ellos una corriente continua casi indefinidamente, sin disminución de intensidad y sin alimentación exterior. No se pierde energía por efecto Joule, y, por tanto, tampoco hay que suministrar energía para mantener la corriente en el superconductor.

Cuando un superconductor se coloca en un campo magnético de intensidad creciente, a cierto valor crítico de la inducción magnética, la resistividad vuelve a aumentar y el superconductor pasa a ser conductor normal. Este campo magnético puede crearse por la corriente que circula por el mismo superconductor, lo que puede suponer un peligro en la práctica, ya que se produciría una liberación súbita de la energía almacenada en el campo magnético. El valor crítico de la inducción magnética depende de la temperatura a la que se encuentra el superconductor.

Los superconductores se clasifican en dos grupos: superconductores del grupo I y superconductores del grupo II. Al grupo I pertenecen ciertas sustancias puras (niobio, vanadio y titanio, por ejemplo), que se caracterizan porque la superconductividad desaparece bruscamente cuando el campo magnético alcanza el valor crítico. Los superconductores del grupo II poseen dos valores de las inducciones magnéticas críticas. Hasta el primer valor estos materiales son superconductores, entre los dos valores de las inducciones críticas su resistividad aumenta paulatinamente y, cuando toman el valor superior de la inducción crítica, se comportan como conductores normales. Los superconductores del grupo II suelen ser compuestos o aleaciones. Este tipo se utilizará en los arrollamientos superconductores en los reactores de fusión.



REFRIGERACION DEL REACTOR. En *a* se esquematiza un reactor de fusión refrigerado por litio natural o por flibe. La refrigeración directa de la capa fértil por litio es interesante por su elevada conductividad térmica y alto calor específico. Los inconvenientes fundamentales de este sistema radican en su notable poder corrosivo y en la facilidad que tiene el litio de reaccionar de manera explosiva con el agua y el aire. Además, por ser fluido conductor, al moverse dentro de campos magnéticos se inducen en su seno corrientes que originan caídas de presión de origen magnetohidrodinámico. La figura *b* corresponde al caso de refrigeración indirecta de la capa fértil, por ejemplo, por medio de un gas inerte como el helio; no existen, por lo tanto, problemas de corrosión, ni de irradiación, ni de interacción con el campo magnético, toda vez que el helio no es conductor. La conversión de energía se realiza en el primer caso por medio de un ciclo convencional de agua y vapor de agua recalentado (ciclo Rankine) y en el segundo mediante un ciclo Brayton en una instalación de turbina de gas, sin que deba por ello descartarse la posibilidad de recurrir al ciclo convencional con turbina de vapor, en cuyo caso debería procederse a la generación de vapor de agua en un intercambiador de calor, como ocurre en el caso *a* (*centro de la ilustración superior*).

Se emplearán como materiales compuestos de titanio y niobio, estaño y niobio, galio y vanadio, aluminio y niobio, entre otros. Estos materiales pueden transportar grandes corrientes, para generar campos magnéticos de más de 10^5 gauss. Sus temperaturas críticas están entre 4 y 20 grados Kelvin, según el material de que se trate. Para ello se les refrigera con helio líquido. Las cantidades de helio necesarias para la refrigeración de los arrollamientos superconductors en el reactor de fusión son muy grandes, de alrededor de 15 toneladas.

Un peligro que encierran los arrollamientos superconductores es que el superconductor pase a ser en ciertas pequeñas zonas un conductor normal, debido a una fluctuación del campo magnético en ese punto, por ejemplo. Dada la intensa corriente que circula por los arrollamientos, ello podría acarrear consecuencias catastróficas: se desarrollaría una gran cantidad de calor y el helio entraría en ebullición, aumentando enormemente su presión. De ahí que sea necesario estabilizar los arrollamientos superconductores contra estos fallos; lo que se consigue a través de la llamada estabilización criogénica, que consiste en rodear al superconductor de algún buen conductor normal y refrigerarlo por helio líquido. En el caso de que el superconductor pasase a ser en zonas concretas un conductor normal, la corriente se desviaría inmediatamente a través del conductor, de cobre o aluminio, y el calor que se desarrollase sería eliminado por ebullición del helio líquido. Se evitaría de esta forma la propagación de la zona normal en el superconductor, y éste podría reestablecerse.

También constituye un problema la descarga de la corriente en los arrollamientos superconductores. La corriente alcanza unos 10 kiloampère, y la energía almacenada en su campo magnético es de alrededor de los 10^5 megajoule. La descarga debe realizarse en un intervalo de algunos minutos, y no de un modo brusco. La energía de la corriente que circula por los arrollamientos superconductores se disipa en una carga resistiva exterior, a través de varios circuitos. Los arrollamientos tienen que hacer frente a las fuertes tensiones que se producen. La principal fuerza que actúa es la dirigida hacia el eje del toro; dicha fuerza resulta de la atracción que ejercen entre sí cada dos segmentos opuestos del toro, al circular por ellos corrientes paralelas y en el mismo sentido. Para soportar estas fuerzas y evitar que los arrollamientos se desplomen hacia el centro, es necesario que en el centro del toro haya una fuerte columna de acero, a modo de soporte.

Son muchos y muy difíciles los pro-

blemas físicos y tecnológicos que han de resolverse antes de que se pueda construir un reactor de fusión. Se ha de profundizar en la estabilidad del confinamiento del plasma y en los métodos de calentamiento, hay que dar soluciones a la inyección de combustible, extracción de impurezas y productos de reacción, protección de la primera pared, etcétera. Deben encontrarse nuevos materiales que resistan las condiciones a las que han de verse sometidos en el reactor de fusión y que no sufran un deterioro demasiado rápido por los neutrones y radiaciones electromagnéticas emitidas por el plasma. No es de esperar, pues, que existan reactores de fusión antes del comienzo del próximo siglo, aunque en el próximo decenio quizás empiecen a trabajar pequeños reactores de fusión experimentales.

La fusión nuclear podría significar, en un tiempo no muy lejano, la solución definitiva a los problemas que padece la humanidad para el abastecimiento de energía. Presenta una serie de ventajas sobre otras fuentes de energía. Se trata de un recurso prácticamente inagotable; sus combustibles –deuterio y litio– son baratos, de fácil extracción y elaboración y están al alcance de todos los países. Puede abastecer cualquier demanda de energía en el futuro. Se trata de una fuente de energía limpia, pues no existen problemas de contaminación química ni de transporte de combustibles radiactivos (el tritio sigue un ciclo cerrado en el reactor), ni problemas de eliminación de desechos radiactivos. La contaminación térmica puede mantenerse en niveles bajos, ya que permite trabajar a grandes temperaturas, alcanzando elevados rendimientos de conversión de energía. En algunos sistemas de confinamiento es posible, además, un proceso de conversión directa de la energía de las partículas cargadas en electricidad.

La fusión nuclear es una fuente de energía segura. No hay peligro de que el reactor se salga de control provocando una catástrofe, ya que antes de que la densidad del plasma alcance un valor desencadenante de explosión se enfría éste. Al no haber desechos radiactivos, el problema del calor residual es menos grave que en otras formas de producción energética. Los mayores riesgos del reactor de fusión están en las fugas o escapes de tritio radiactivo, rotura de alguna conducción de litio y reacción explosiva del mismo con el aire de la atmósfera, liberación incontrolada de la energía almacenada en el campo magnético y radiactividad inducida.

El curioso espectro de SS 433

El corrimiento de las líneas de emisión hallado en el espectro de este objeto estelar sugiere que está expulsando dos finos chorros de materia a gran velocidad en direcciones opuestas. ¿Cómo explicar tal comportamiento?

Bruce Margon

Con 100.000 millones de estrellas en la Vía Láctea no hay mucho margen para individualidades. Virtualmente, todos los fenómenos de la galaxia tienen probabilidad de ocurrir más de una vez. De aquí que "único" no es una palabra de aplicación frecuente en astronomía estelar. Incluso a algunos objetos astronómicos muy fuera de lo corriente se les ha encontrado análogos; por ejemplo, la Nebulosa del Cangrejo, espectacular resto de una explosión de supernova que se observó en el año 1054, tiene al parecer dos o tres contrapartidas en otros lugares de la galaxia. Sin embargo, en los dos últimos años se han acumulado observaciones sobre un objeto estelar que probablemente es merecedor del término único. El objeto, conocido como SS 433, parece estar expulsando dos chorros de materia muy delgados en direcciones opuestas a velocidades increíblemente altas. Nunca se supo de otra estrella que se comportara así.

SS 433 es un objeto fascinante por muchas razones, de las que no es la menor el hecho de haber sido descubierto independientemente en varias ocasiones en los dos decenios últimos, aunque cada vez pasó desapercibido a sus descubridores su carácter verdaderamente exótico. El objeto emite una excepcional distribución de radiación en diversas partes del espectro electromagnético (en longitudes de onda correspondientes a luz visible, ondas de radio y rayos X), y cada una de esas peculiaridades había sido observada individualmente a lo largo de los años. Pero es muy reciente la aparición de un cuadro completo de SS 433. Es instructivo seguir la primitiva historia de las observaciones para ver cuán cerca estuvieron los primeros investigadores de detectar las propiedades únicas del objeto.

Las primeras indicaciones pertinentes se obtuvieron a principios del decenio de 1960, cuando los astrónomos de la Universidad Case Western Reserve llevaron a cabo una exploración sistemática dirigida al descubrimiento de ciertos tipos de débiles estrellas cerca del plano central de nuestra galaxia. La técnica telescópica particular que emplearon incluía un objetivo prismático, elemento óptico dispersor de la luz que alarga las imágenes de las estrellas en la placa fotográfica, convirtiéndolas en débiles trazas que representan el espectro visible de la luz de la estrella. A diferencia de las técnicas más convencionales de astronomía espectroscópica, en las que la luz procedente de una sola estrella se hace pasar a través de una estrecha rendija cerca del foco del telescopio y se dispersa después sobre una parte considerable de la placa fotográfica, la técnica del objetivo prismático sólo separa ligeramente las longitudes de onda de la luz de la estrella. Por otra parte, lo hace simultáneamente para toda imagen estelar visible en el campo del objetivo, en lugar de hacerlo cada vez para una estrella. De este modo, se obtienen unos espectrogramas muy toscos, apropiados principalmente para la clasificación de distintos tipos de objetos, a partir de un gran número de estrellas y con sólo un consumo moderado del precioso tiempo de utilización del telescopio.

La exploración Case se proyectó específicamente para hallar nuevas estrellas con líneas de emisión en sus espectros. La existencia de tales líneas forma la base de una técnica de diagnóstico de amplia aplicación, común a muchas áreas de la física y la química. Cuando los electrones ligados a los átomos de un gas se excitan a estados orbitales de más alta energía, bien por medio de choques con otras partículas o de interacciones radiativas, su subsiguiente desexcitación crea nuevos fotones, o cuantos de luz, a ciertas longitudes de onda discretas que están determinadas por la diferencia de energía entre los estados orbitales, y por tanto, por la estructura electrónica del elemento de que se trate. La aparición de esas líneas (llamadas así porque aparecen como una raya oscura a cada longitud de onda particular en el negativo fotográfico expuesto) en un espectrograma puede servir, pues, como un detector muy específico de las condiciones físicas del gas emisor. La longitud de onda precisa de tal línea, por ejemplo, indica sin ambigüedad cuál es el elemento químico original emisor de la luz, ya que lleva la memoria de la estructura electrónica del átomo.

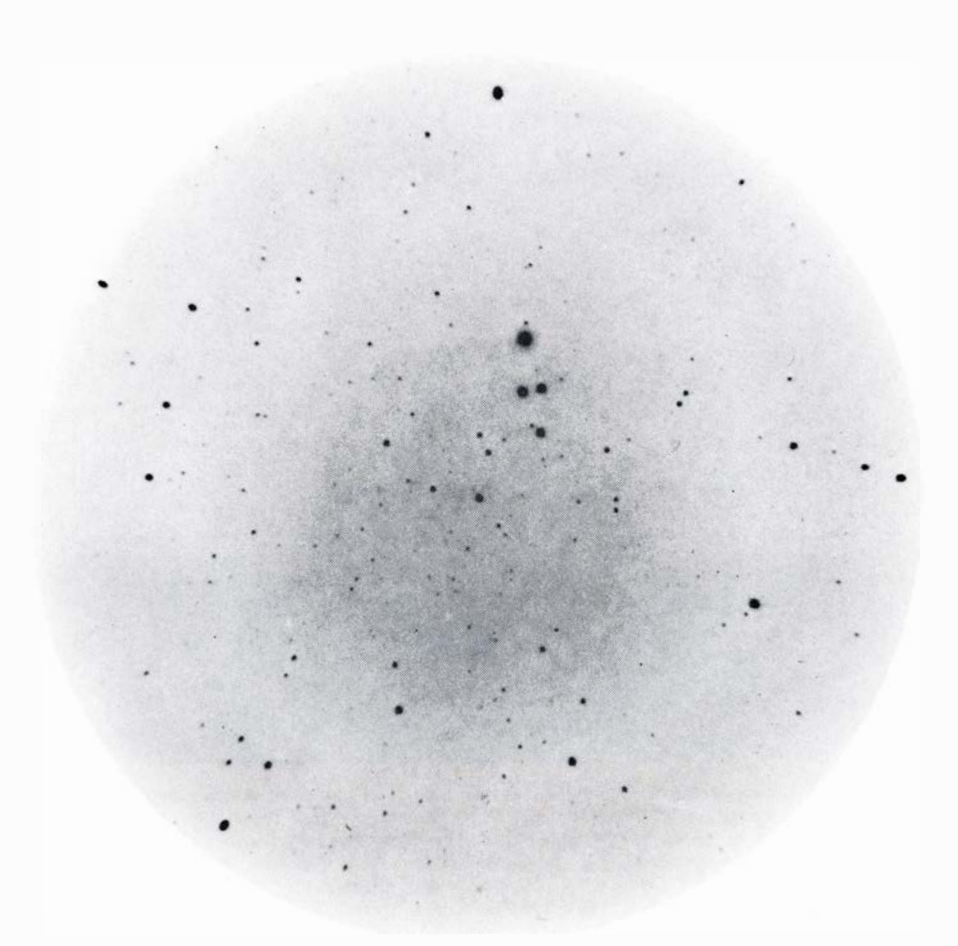
Estudios mediante objetivos prismáticos tales como el llevado a cabo por los investigadores de Case han demostrado que tal vez un alto porcentaje, del orden del 10 por ciento, de todas las estrellas produce líneas de emisión características de un gas caliente y excitado; se cree que tal gas existe en una capa exterior y difusa que rodea a una estrella por lo demás corriente. Hay muchas posibles razones para esta situación anormal, y varían de una estrella a otra. Por ejemplo, se encuentran frecuentemente líneas de emisión asociadas con estrellas muy jóvenes y muy antiguas. En ocasiones también se hallan identificadas con estrellas estables de edad intermedia.

Una de las placas expuestas por los investigadores de Case estaba centrada en la constelación Aguila, en el centro de la Vía Láctea; en la placa se veían muchos objetos emisores de intensas líneas. En 1977, C. Bruce Stephenson y Nicholas Sanduleak publicaron una lista de objetos emisores de líneas en el área; el número 433 de su lista era la estrella ahora conocida por SS 433. Stephenson y Sanduleak no tenían motivos para individualizar SS 433 como algo particularmente distinto de los otros cientos de estrellas emisoras de líneas identificadas en su exploración; para eso se hubieran necesitado datos

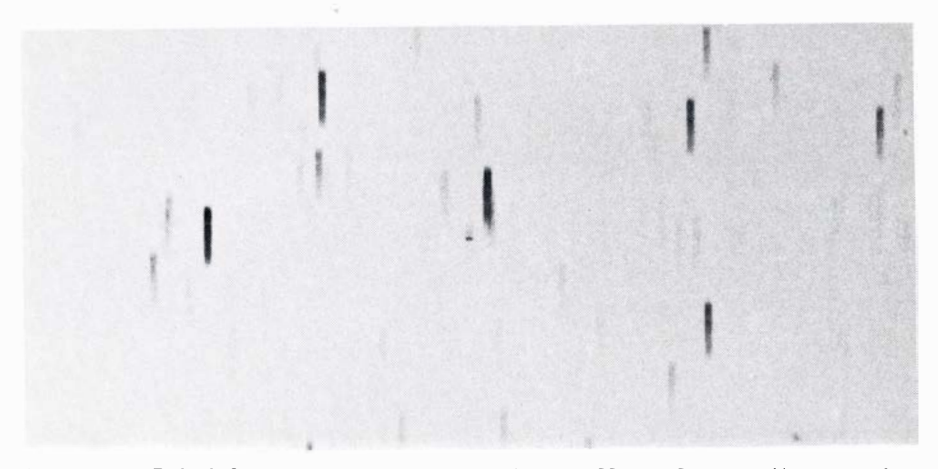
espectroscópicos más detallados que los que podían obtenerse mediante la técnica del objetivo prismático. De hecho, uno de los principales objetivos de exploraciones como la de Case es estimular en el futuro observaciones detalladas de objetos de aspecto interesante. El resultado logrado por Stephenson y Sanduleak a este respecto en el caso de SS 433 es muy superior a sus más fantásticos sueños.

Con un brillo aparente que equivale al de una estrella de decimocuarta magnitud, SS 433 es más de 1000 veces inferior al valor umbral que permitiría su observación a simple vista. Se trata, sin embargo, de un blanco fácil para observaciones espectroscópicas, incluso con un telescopio pequeño. Las subsiguientes observaciones pudieron arrancar muy bien de la temprana fecha de 1975, cuando Lawrence Krumenaker, otro astrónomo de Case, escribió un breve artículo que apareció antes de la publicación del catálogo Stephenson-Sanduleak. El artículo mencionaba un pequeño subconjunto de estrellas emisoras de líneas contenido en el catálogo completo, entre las que se encontraba la entonces anónima SS 433 y daba las coordenadas celestes y los mapas necesarios para que otros astrónomos pudieran localizar esos objetos. Por desgracia, las coordenadas que se daban para SS 433 eran incorrectas, probablemente a causa de un simple error de transcripción. El objeto no se podría haber hallado de nuevo aunque hubiera picado la curiosidad de otro observador.

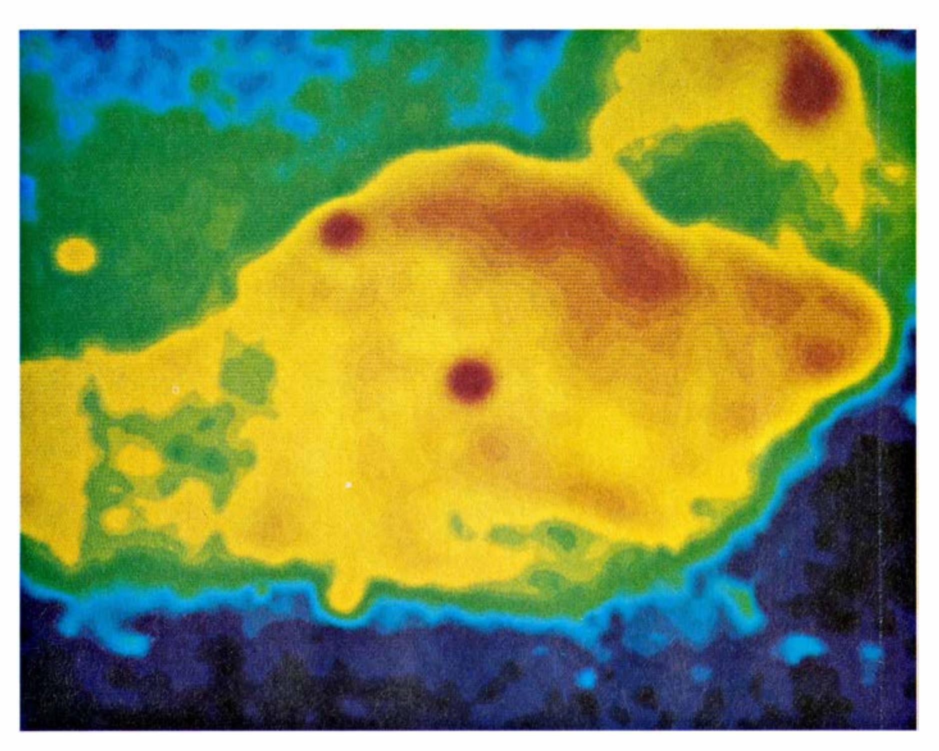
Aproximadamente hacia la misma época en que la exploración de Case se llevaba a cabo, un programa de observación sin relación alguna con él descubría también, sin pretenderlo, la SS 433. Los radioastrónomos de la Universidad de Cambridge estaban compilando una de sus varias listas exhaustivas de radiofuentes celestes, en este caso la lista que saldría como Cuarto Catálogo de Cambridge. La mayoría de las fuentes de radioondas incluidas en el catálogo resultan estar asociadas, bajo examen detallado de fotografías ópticas, con distantes objetos extragalácticos tales como galaxias y quasars. La radioemisión de estrellas normales en nuestra galaxia es demasiado débil para poder detectarse en tales exploraciones, incluso para objetos próximos al sistema solar. En el curso de la exploración, los investigadores de Cambridge encontraron una radiofuente bastante brillante en Aguila y la designaron 4C 04.66.



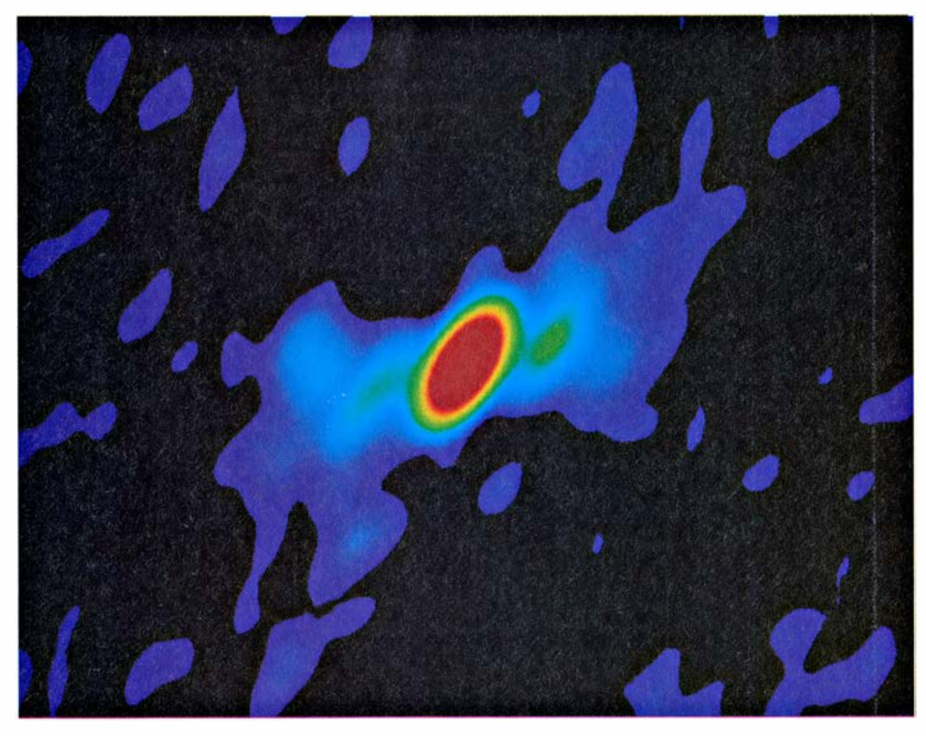
SS 433 APARECE MUY POCO DESTACADO en las fotografías astronómicas corrientes, como ésta, hecha por Eugene A. Harlan con el telescopio refractor de 90 centímetros en el Observatorio Lick de la Universidad de California. La imagen de SS 433 es el punto negro en el centro exacto de este negativo fotográfico circular, que abarca un campo visual de aproximadamente una décima de grado de diámetro en la constelación del Aguila. Clasificado en términos de su brillo aparente como un objeto de decimocuarta magnitud, habría que multiplicar el resplandor de SS 433 por 1000 para que pudiera observársele a simple vista. La evidente carencia de cualquier peculiaridad visible que distinga el objeto de sus numerosas estrellas vecinas puede explicar por qué se tardó tanto en su descubrimiento.



PRIMERA INDICACION de la naturaleza poco corriente de SS 433. Se descubrió en esta placa espectroscópica de campo relativamente grande obtenida hace más de 20 años en el curso de una exploración estelar especial llevada a cabo por los astrónomos del Observatorio Warner y Swasey, de la Universidad Case Western Reserve. La placa se obtuvo mediante la técnica del objetivo prismático, que produce el efecto de difuminar las imágenes estelares convirtiéndolas en cortos trazos que representan el espectro visible de la luz de la estrella. La exposición en este caso se limitó a la parte roja del espectro, mediante el uso de filtros apropiados, incluyendo longitudes de onda en el intervalo entre 6000 y 6800 unidades ångstrom. En la porción del cielo cubierta por esta placa particular (una superficie de aproximadamente medio grado de diámetro, casi igual al diámetro de la luna llena) sólo hay una imagen espectral que aparece en forma de línea en lugar de traza alargada; es la anodina raya cerca del centro de la placa. La mayor parte de la luz roja procedente de este objeto está concentrada en la línea de emisión alfa del hidrógeno, muy fuerte, en vez de hallarse distribuida de una forma continua a lo largo de todas las longitudes de onda de este intervalo, como es el caso de las estrellas vecinas. En 1977, C. Bruce Stephenson y Nicholas Sanduleak, de la Case Western Reserve, publicaron una lista de los objetos que presentaban tales brillantes líneas de emisión en esta parte del cielo. El objeto del centro de la placa llevaba el número 433 en la lista de Stephenson-Sanduleak, motivo por el que se le designó SS 433.



RADIOMAPA de una gran región que rodea SS 433, que revela que este objeto está inmerso en una extensa fuente de radioemisión, conocida como W 50, la cual se supone ser el resto de una antigua explosión de supernova. El mapa, que cubre una región del cielo de aproximadamente dos grados de diámetro, se presenta en una clave de colores: el rojo representa la radioemisión más intensa y el azul la menos intensa. El punto rojo brillante cerca del centro coincide con SS 433. La mancha semejante hacia el Nordeste (*parte superior izquierda*) es probablemente una radiofuente extragaláctica compacta que se ve a través del borde de W 50. La radiofuente extensa hacia el Noroeste (*parte superior derecha*) es una nube de hidrógeno ionizado sin relación con W 50 y situada entre ésta y el sistema solar. Los datos fueron registrados en longitud de onda de 11 centímetros por B. J. Geldzahler, T. Pauls y C. J. Salter, trabajando con el radiotelescopio de 100 metros del Instituto Max Planck de Radioastronomía.



RADIOMAPA DE ALTA RESOLUCION de la región en la inmensa vecindad de SS 433, realizado a una longitud de onda de seis centímetros con la Gran Distribución, un complejo de radiotelescopios cerca de Socorro, Nuevo México. Las dos estructuras alargadas que emanan de la fuente central puntual en el mapa están alineadas con los lóbulos en la radiofuente circundante, W 50. El mapa es fruto de un estudio en colaboración llevado a cabo por John T. Stocke, de la Universidad de Arizona, y Ernest R. Seaquist y William S. Gilmore, de la Universidad de Toronto. La representación en color del mapa fue preparada por Eric W. Greisen, del Observatorio Nacional de Astronomía norteamericano. La elipticidad de la fuente central en esta representación es un efecto del proceso de construcción de la imagen.

Si al llegar a este punto alguien hubiera notado la coincidencia de la posición de la radiofuente con la posición de una estrella visible bastante brillante, es casi seguro que se habrían emprendido desde entonces nuevas observaciones ópticas, habida cuenta de la escasez de radiofuentes detectables visualmente. Pero una vez más, las coordenadas celestes publicadas para 4C 04.66 no coincidían con la posición real de SS 433. En este caso, la razón probable para la discrepancia era instrumental; esa región particular del cielo es una confusa maraña de radiofuentes principales y de fondo, y los instrumentos usados en la exploración de Cambridge eran incapaces de fijar con exactitud la posición de una fuente, por más intensa que fuera, frente a tal confusión. De nuevo se perdió la oportunidad de observar la especial naturaleza de SS 433.

En realidad, parte de la razón para la confusión en este caso está relacionada directamente con la historia de SS 433. Centrada cerca de esa región del cielo hay una extensa e intensa mancha de radioemisión que cubre un área doble del tamaño angular de la luna llena. La zona radioemisora es también un objeto que se había catalogado antes y estaba casi olvidado, el W 50. (La designación deriva de la aparición de la fuente en un catálogo de imágenes de radio compilado en el decenio de 1950 por el astrónomo Gart Westerhout, de la Universidad de Maryland.) La radioestructura y el radioespectro de W 50 han llevado a la mayoría de los investigadores a sugerir que es un viejo resto de supernova: los difusos residuos en expansión de una antigua explosión estelar semejante a la Nebulosa del Cangrejo. De no haber registro histórico indicador de la observación de una supernova en esa parte del cielo, ha resultado difícil obtener una prueba absoluta de esta afirmación acerca de la naturaleza de W 50. En la hipótesis de que se tratara realmente de un resto de supernova, su tamaño y brillo radiante, cuando se compara con datos análogos de restos de supernova mejor estudiados, sugiere que la explosión tuvo lugar hace unos 10.000 años. Aun cuando tal edad resulta a todas luces excesiva para que la explosión hubiera sido registrada por observadores humanos, todavía clasificaría el suceso que produjo W 50 como una de las supernovas más recientes.

En 1975, David H. Clark y sus colegas de la Universidad de Sidney publicaron una serie de radioobservaciones proyectadas para cartografiar la

estructura detallada de W 50. El mapa muestra también la fuente brillante y casi puntual que ahora se sabe es SS 433, rodeada por la difusa emisión de W 50. Se produjo entonces otro hecho casual más que contribuyó a bloquear el ulterior estudio de esta curiosa configuración. En efecto, aunque la radioestructura real de W 50 es bastante simétrica, con SS 433 cercano al centro, el mapa publicado por Clark y sus colaboradores mostraba sólo la mitad Norte del resto, lo que en cierta manera enmascaraba la prominente posición central de la misteriosa fuente puntual. Si otros observadores se hubieran dado cuenta de la sorprendente simetría del resto y hubieran observado, por consiguiente, que la radiofuente puntual no identificada estaba centrada de forma muy precisa, es casi seguro que ello habría estimulado nuevas observaciones.



INTENSA RADIOFUENTE coincidente con la posición de SS 433, que aparece en el centro de esta imagen generada por ordenador a partir de los datos suministrados por un telescopio de rayos X a bordo del satélite *HEAO-2*, conocido también como Observatorio Einstein de Rayos X. La débil indicación de emisión de rayos X que se extiende hacia fuera, a partir de la fuente central y a ambos lados, sugiere la presencia de chorros de gas caliente que se dirigen hacia los extremos Este y Oeste de W 50. Esta imagen de rayos X, producto de una exposición de cinco horas, fue realizada pro Seaquist y Gilmore con Jonathan E. Grindlay y Frederic D. Seward, del Centro de Astrofísica del Harvard College Observatory y del Observatorio Astrofísico Smithsoniano. El campo de visión es aproximadamente de un grado.

A pesar de la creencia generalizada de que las supernovas dejan atrás frecuentemente, tal vez siempre, un resto estelar colapsado y extraño, una estrella de neutrones o un agujero negro por ejemplo, sólo hay dos casos en que se ha encontrado sin ambigüedad una estrella colapsada en un resto de supernova. Los restos, visible y de radio, tanto del suceso del Cangrejo como de otro en la constelación Vela alojan también un pulsar, lo que sugiere la existencia de una estrella de neutrones en rotación. El que únicamente se conozcan dos de tales coincidencias, a pesar de la existencia de docenas de restos de supernova emisores de luz y de ondas de radio y de cientos de pulsares, es molesto aunque quizás no tenga mucha trascendencia. A buen seguro, no se habría perdido la oportunidad de explorar otra coincidencia potencial de este tipo si se hubiera reconocido a tiempo.

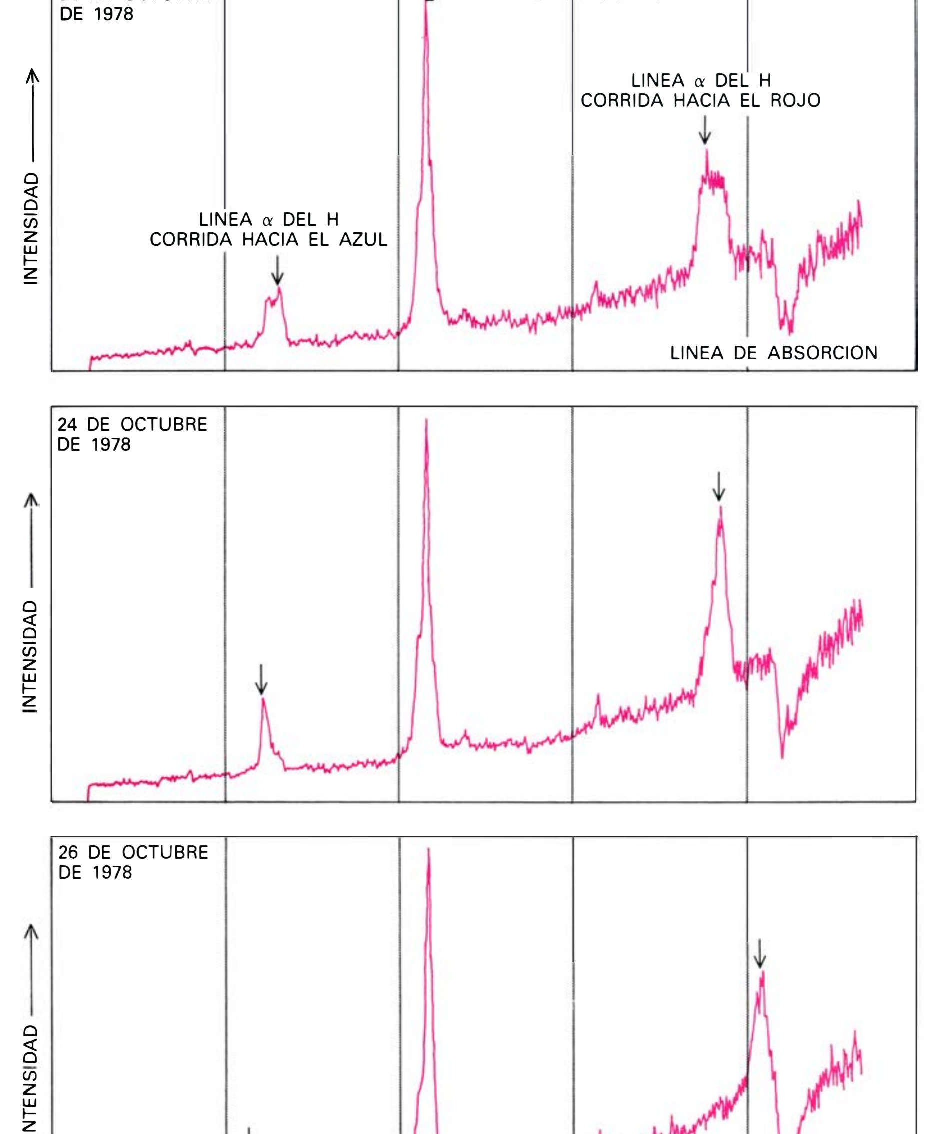
El prefacio final para el reconocimiento de las extrañas propiedades de SS 433 se refiere a su emisión de rayos X. Como ocurre con la radioemisión, muy pocas estrellas normales son fuentes detectables de rayos X, aunque de nuevo hallamos aquí que muchos restos de supernova son fuentes de rayos X. A principios del decenio de 1970, dos satélites en órbita alrededor de la Tierra registraron independientemente emisiones de rayos X procedentes de la vecindad del Aguila; de nuevo, la fuente de esas emisiones, que ahora se sabe ser SS 433, fue debidamente denominada y catalogada. Los observadores que trabajaban con los datos del satélite norteamericano *Uhuru* designaron a la fuente de rayos X 4U1908 + 05 (del cuarto catálogo *Uhuru*, con las coordenadas celestes aproximadas de la posición en el cielo). Contemporáneamente, un grupo de predominio británico (dirigido por un investigador norteamericano visitante, Frederick D. Seward), trabajando con los datos del satélite británico *Ariel V,* clasificaba al objeto como el A1909 + 04.

La baja resolución espacial de los experimentos impidieron que cada uno de los grupos apreciara la coincidencia de la fuente de rayos X y la brillante estrella. Seward y sus colaboradores notaron, sin embargo, que la intensidad de la emisión de rayos X parecía variar con el tiempo, característica no observada en la emisión de rayos X procedente de restos de supernova en virtud de su evolución, lenta en comparación, tras la explosión inicial. Seward y sus colaboradores comentaron con sorprendente intuición en su informe de 1976 que probablemente la fuente de rayos X no era sólo el resto W 50, sino quizás algo más extraño relacionado con él.

Por último, en el verano de 1978 emergió una síntesis de estos numerosos indicios de las extrañas características de SS 433, como resultado de los esfuerzos contemporáneos, si bien en gran parte independientes, de tres grupos distintos de investigadores. El grupo de radioastrónomos canadienses, dirigido por Ernest R. Seaquist, de la Universidad de Toronto, investigaba las emisiones procedentes de estrellas jóvenes, llevando a cabo una nueva exploración de objetos registrados en el catálogo Stephenson-Sanduleak. Aunque esto condujo a todavía un nuevo redescubrimiento de la radiofuente, Seaquist y sus colegas apreciaron correctamente que la fuente estaba asociada al brillante objeto estelar SS 433. Al mismo tiempo, otro grupo de radioastrónomos, en este caso británicos y australianos bajo la dirección de Sir Martin Ryle, de Cambridge, habían emprendido el estudio del problema, completamente diferente, de la escasez observada de radiofuentes puntuales en el interior de los extensos restos de supernovas. En una afinada exploración cuyo objetivo era hallar y precisar las posiciones de tales objetos, de nuevo hallaron la radiofuente en W 50

y también observaron la coincidencia con el objeto de magnitud óptica 14.

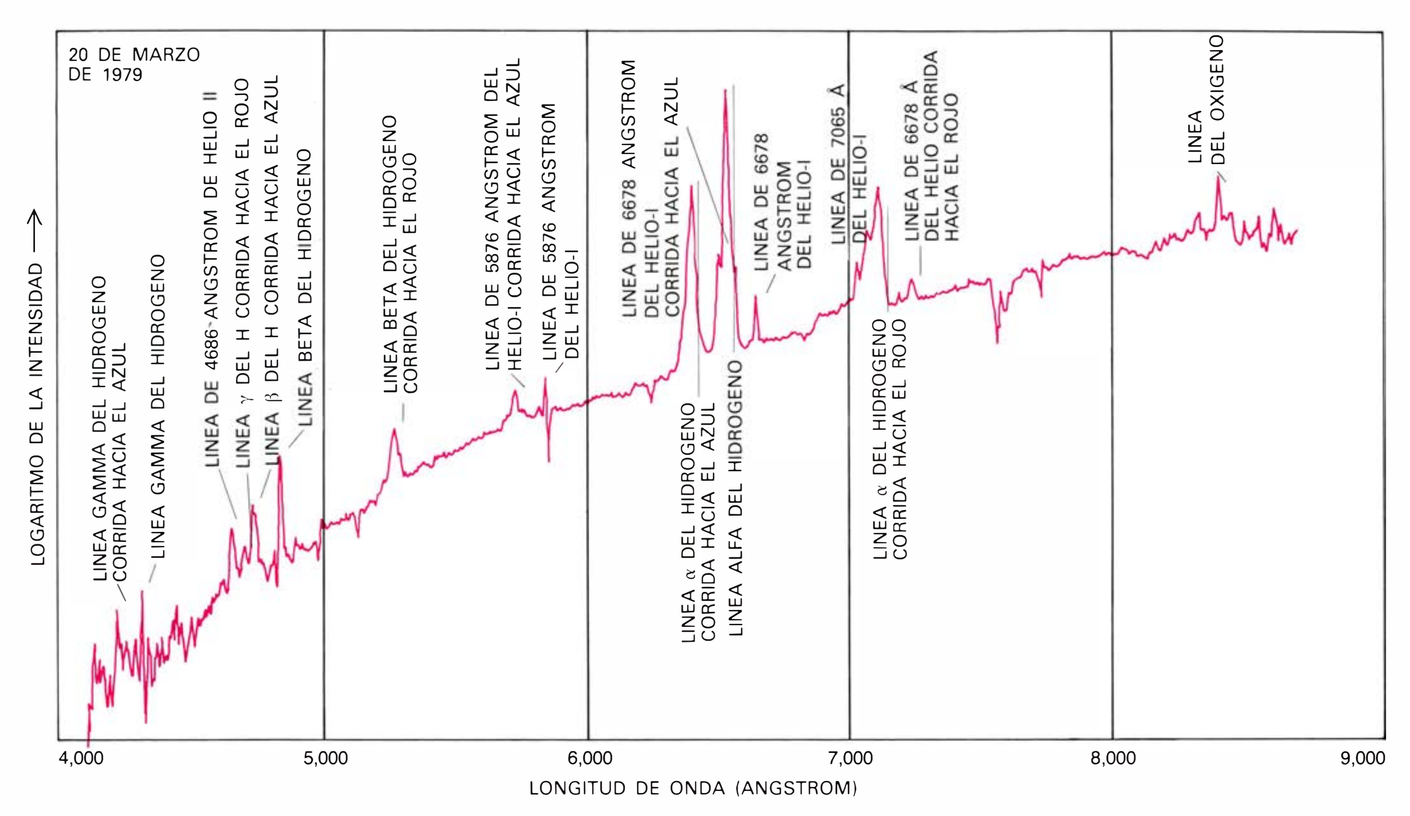
Finalmente, Clark y su colega Paul Murdin decidieron obtener espectrogramas de estrellas visibles cuya posición fuese próxima a la de la radiofuente, bastante defectuosamente situada, que Clark había previamente anotado en su mapa de W 50. Trabajando con el Telescopio Angloaustraliano en Australia, registraron el espectro de SS 433 en junio de 1978; eran las primeras observaciones espectroscópicas publicadas desde que se habían tomado las toscas placas con objetivo prismático en el catálogo de Stephenson-Sanduleak. Clark y Murdin no se dieron cuenta de inmediato de que estaban observando un objeto ya catalogado. Sin embargo, el espectrograma que obtuvieron dejaba pocas dudas de que habían localizado correctamente la contrapartida visible de la extraña radiofuente. Sus observaciones mostraban líneas de emisión de una intensidad que se encuentra sólo en las estrellas menos corrientes. Una radioposición extremadamente exacta suministrada por Ryle y su grupo hizo concluyente la identificación de SS 433 con la radiofuente; las posiciones de los objetos radio y óptico coincidían de manera precisa. La conexión final e irrevocable con la fuente de rayos X se forjó en una reciente serie de medidas de rayos X, de suma exactitud, llevada a cabo por Seaquist y su grupo con el satélite *HEAO-2* en órbita alrededor de la Tierra, llamado también Observatorio Einstein de rayos X; se sabe ahora que la posición de rayos X concuerda perfectamente con las posiciones óptica y de radio.



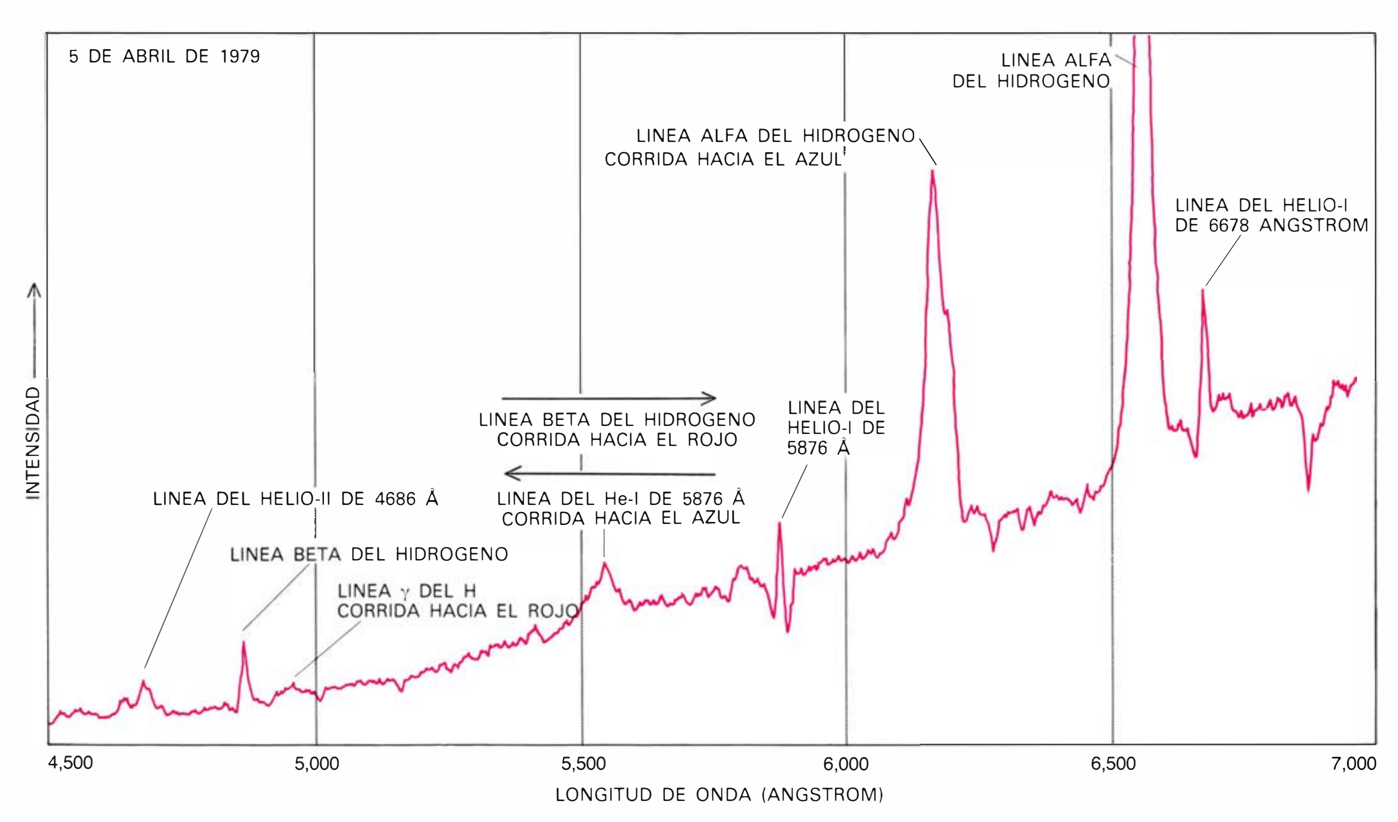
ESPECTRO DE SS 433, registrado en tres noches diferentes en un período de cuatro días por Remington P. S. Stone trabajando con el telescopio reflector de 60 centímetros del Observatorio Lick. La característica más prominente, el máximo a la longitud de onda de 6536 ångstrom, corresponde a la extremadamente fuerte línea de emisión alfa del hidrógeno. Se pueden detectar líneas de emisión mucho más débiles, atribuibles al helio, a las longitudes de onda de 5876, 6678 y 7065 ångstrom. A ambos flancos de la línea central pueden verse dos líneas de emisión muy fuertes, que ahora se sabe son versiones de la línea alfa central del hidrógeno corridas por efecto Doppler; una de ellas se encuentra hacia el extremo del espectro correspondiente a longitudes de onda más cortas (azul) (*a la izquierda*), y la otra hacia el extremo de mayores longitudes de onda (rojo) (*a la derecha*). En el curso de las tres noches de observación, la línea corrida hacia el rojo se movió claramente más hacia el rojo, y la corrida hacia el azul se movió más hacia el azul. La fuerte caída en las curvas hacia la longitud de onda de 7600 ångstrom no tiene relación con SS 433; es una línea de absorción causada por moléculas de la atmósfera terrestre.

Clark y Murdin publicaron sus observaciones en una breve nota en *Nature* en el otoño de 1978. Identificaron las prominentes líneas de emisión como longitudes de onda apropiadas para los átomos excitados de hidrógeno y helio (lo que no es sorprendente en sí mismo, ya que esos son invariablemente los elementos químicos más abundantes en las estrellas). También aludían indirectamente a líneas de emisión más débiles de origen incierto en el espectro. (Esas características mencionadas de pasada demostraron más tarde constituir la clave de la totalidad del misterio de SS 433.) Por último, destacaron la sorprendente triple coincidencia de una estrella emisora de líneas en el visible, una fuente de rayos X y una radiofuente, las tres centradas en un resto de supernova; y sugirieron que SS 433 y W 50 podían estar causalmente relacionadas.

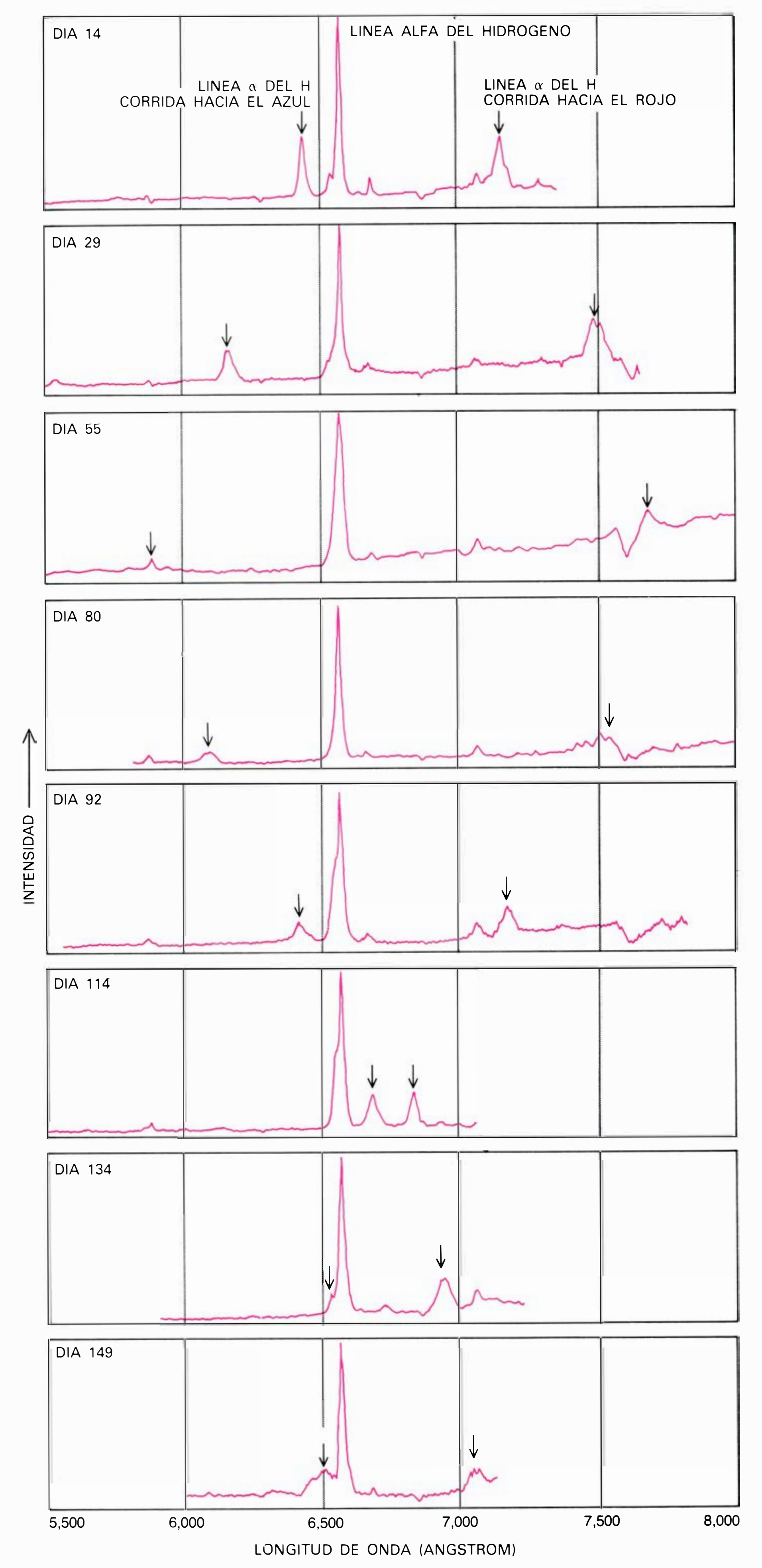
Ahí comenzó mi propia relación con SS 433. Me había venido interesando, desde hacía tiempo, por las características ópticas de las débiles contrapartidas visibles de las fuentes celestes de rayos X. La mayoría de las estrellas de rayos X en nuestra galaxia resultan estar en sistemas binarios que constan de un objeto compacto, tal como una enana blanca o una estrella de neutrones, ligado por una órbita pequeña a una estrella comparativamente normal, con frecuencia no demasiado diferente del Sol. La atracción gravitatoria ejercida por la estrella compacta sobre su compañera es causa habitual de que la estrella, por lo demás normal, pierda



LAS LINEAS DE EMISION TRIPLICADAS en este espectrograma particular de SS 433 fueron lo que convenció al autor y sus colegas de la Universidad de California en Los Angeles de que las líneas "móviles" de emisión observadas en varios espectros de este objeto deben atribuirse al efecto Doppler. El espectrograma fue obtenido por Steven A. Grandi, de la Universidad de California en los Angeles, en la noche del 20 de marzo de 1979 con el telescopio reflector Shane de tres metros del Observatorio Lick. En el espectro, cierto número de prominentes líneas de emisión del hidrógeno y el helio aparecen cada una representada tres veces: por una fuerte componente central flanqueada por dos componentes corridas por efecto Doppler, una de ellas hacia el rojo y la otra hacia el azul. Las fuertes bandas de absorción que se aprecian en el espectro carecen de relación con el objeto estelar.



ALGUNAS VECES, el gran número de líneas de emisión "móviles" producidas por efecto Doppler que aparecen en el espectro de SS 433 muestran una coincidencia desconcertante. En este caso, por ejemplo, se identifican varias líneas prominentes de emisión, estacionarias y móviles; entre ellas, el máximo, comparativamente ancho, en el centro (hacia los 5500 ångstrom), que igualmente está compuesto de una línea de helio corrida hacia el azul moviéndose hacia la izquierda y una línea de hidrógeno corrida hacia el rojo moviéndose hacia la derecha. La noche en que se obtuvo este espectro, las dos líneas estaban cruzándose al moverse en direcciones opuestas. El espectro se registró en el Lick por Lawrence H. Aller y Charles D. Keyes.



parte de su masa transfiriéndola al objeto compacto. En el proceso de intercambio, la materia en tránsito suele alcanzar temperaturas y densidades suficientes para producir una copiosa emisión de rayos X. Los estudios espectroscópicos de la estrella normal en tales sistemas suministran con frecuencia valiosos datos sobre las características de ambos objetos en ellos. Las densidades enormemente grandes de la materia en objetos compactos, tales como las estrellas de neutrones, no se pueden alcanzar en laboratorios terrestres; son los datos astrofísicos los que proporcionan virtualmente la única información directa sobre el comportamiento de la materia en esos estados exóticos.

Conseguí mi primer espectrograma de SS 433 en septiembre de 1978, trabajando con el telescopio Shane de tres metros en el Observatorio Lick de la Universidad de California. Me había propuesto simplemente confirmar los resultados de Clark y Murdin. Los datos registrados electrónicamente revelaban las fuertes líneas de emisión del hidrógeno y el helio observadas por esos investigadores y también por Stephenson y Sanduleak. Con gran sorpresa mía, sin embargo, descubrí en el espectro ciertas líneas de emisión muy prominentes, que no me eran familiares. La situación resultaba inquietante. Las líneas espectrales que aparecen en un espectro estelar dependen de la abundancia, la temperatura y la densidad de los distintos elementos de la estrella. En situaciones astrofísicas existe alguna variación en esos parámetros, pero no es infinita; ello determina que el espectroscopista estelar se acostumbre al aspecto de ciertas líneas espectrales familiares. Dar con líneas espectrales de emisión a longitudes de onda completamente arbitrarias es una experiencia que podríamos comparar a la sensación del conductor habituado a una carretera en la que, de repente, se encuentra con nuevos ramales.

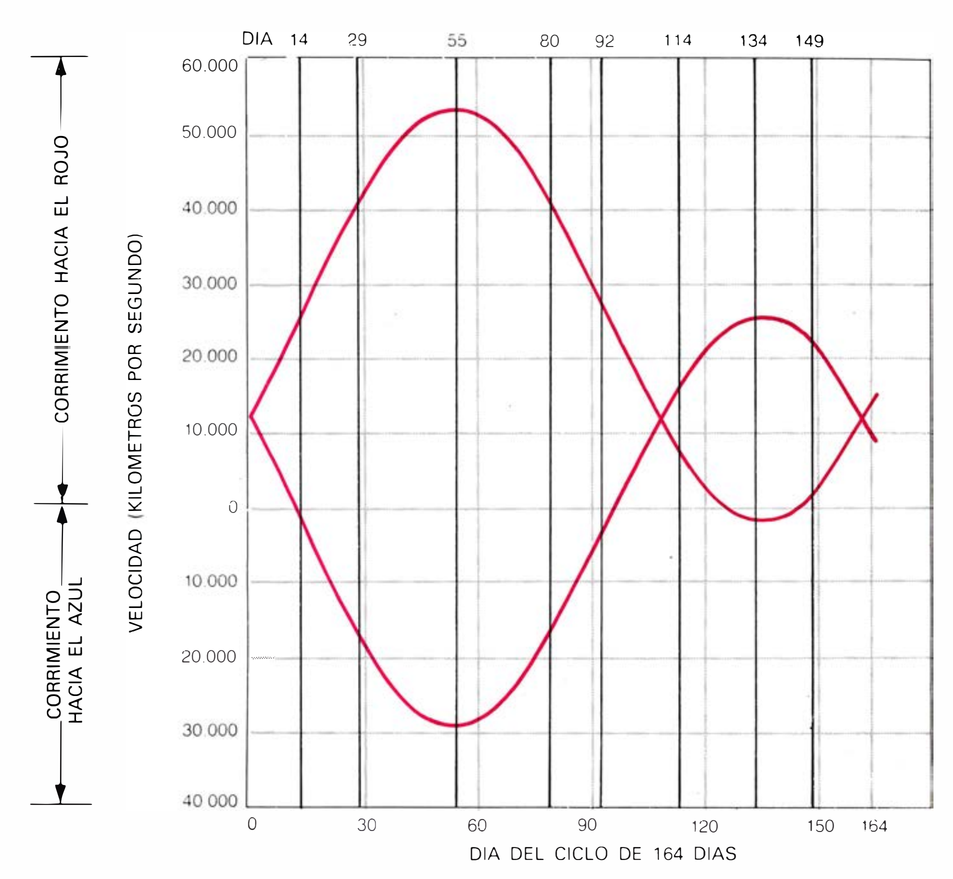
Las intensidades de las líneas de emisión no identificadas en el espectro de SS 433 tenían un interés particular. Estas líneas son sólo algo menos prominentes que las familiares líneas del hidrógeno, que también aparecen en el espectro. Se podría por tanto atribuirlas a un elemento cósmico con una

ESTOS ESPECTROS seleccionados de SS 433, obtenidos por el autor y sus colaboradores a lo largo de un período de 164 días, cubren el ciclo completo del movimiento de las líneas de emisión afectadas por efecto Doppler. Los días en que se hicieron las observaciones se dan, según un modelo del ciclo de 164 días, en la ilustración siguiente.

abundancia comparable a la del hidrógeno. Sin embargo, como se supone que el 90 por ciento de la totalidad de los átomos en la mayoría de las estrellas es de hidrógeno, no existe elemento alguno de abundancia parangonable. ¿Podrían las líneas no identificadas corresponder también al hidrógeno, y podrían estar por alguna razón corridas respecto a las longitudes de onda normales que aparecen en las otras situaciones astrofísicas y de laboratorio?

Existe ese tal mecanismo de corrimiento familiar al espectroscopista: el efecto Doppler. Se sabe que el movimiento relativo de la fuente respecto al observador produce un ligero corrimiento en la longitud de onda observada en cualquier fenómeno ondulatorio, el sonido o la luz por ejemplo. Además, la magnitud del corrimiento refleja en forma conveniente la velocidad del movimiento, y el sentido del corrimiento (hacia longitudes de onda mayores o menores) indica si el objeto se está alejando o acercando. La mayoría de las estrellas tienen un movimiento caótico de unas pocas decenas de kilómetros por segundo con respecto al Sol, lo que origina un corrimiento de las líneas espectrales de aproximadamente 0,01 por ciento respecto a los valores normales. En el caso de algunas estrellas binarias, el movimiento periódico orbital puede ser 10 veces mayor que aquel, causando un corrimiento Doppler proporcionalmente mayor.

Las líneas no identificadas de SS 433, sin embargo, se encontraban en longitudes de onda que no eran de ninguna manera próximas a las de cualquier línea de hidrógeno; por tanto, si se trataba de emisiones de hidrógeno corridas por efecto Doppler, las velocidades necesarias habrían de ser prodigiosas. Para la más prominente de las líneas no identificadas, situada en la parte roja del espectro próxima a la longitud de onda de 7400 angstrom, la velocidad necesaria para dar lugar al corrimiento desde la línea más próxima del hidrógeno (a 6536 angstrom) es de unos 40.000 kilómetros por segundo, ¡más del 10 por ciento de la velocidad de la luz! Como la velocidad de escape de la galaxia sólo es de algunos cientos de kilómetros por segundo, nunca se encuentran velocidades estelares mayores que esas; un tal objeto abandonaría rápida (en la escala de tiempos astronómicos) y absolutamente la galaxia. Resumiendo, la hipótesis de emisiones por líneas de hidrógeno corridas por efecto Doppler no parecía ser una explicación sólida de las observaciones.



ESTAS CURVAS TEORICAS delinean un ciclo completo de 164 días en la evolución pronosticada de líneas espectrales de emisión corridas hacia el rojo y hacia el azul en SS 433, en la hipótesis de que el gas emisor de luz esté concentrado en dos chorros giratorios y directamente opuestos, cada uno con una velocidad de expulsión de 78.000 kilómetros por segundo o aproximadamente un cuarto de la velocidad de la luz. Las líneas negras indican los días en que se obtuvieron los espectros seleccionados que aparecen en la página opuesta. Las líneas de emisión en el espectro concuerda muy bien con los corrimientos hacia el rojo y hacia el azul predichos. Las curvas muestran un corrimiento medio hacia el rojo constante, equivalente a una velocidad de expulsión de 12.000 kilómetros por segundo, que resultan del efecto llamado de dilatación del tiempo en relatividad especial. (Ilustración realizada por Allan Beechel.)

Una característica todavía más extraña del espectro de SS 433 se hizo patente tras varias noches de repetido estudio del objeto. Se observó que las rayas no identificadas a longitudes de onda no corrientes cambiaban de longitud de onda y en cuantía muy apreciable. Así, en un período de cuatro noches, la intensa traza en la parte roja del espectro aumentó su longitud de onda en más de un 1 por cien. Si el efecto Doppler era responsable de ello, esta discrepancia aparentemente pequeña necesitaría explicarse mediante un cambio de velocidad de casi 5000 kilómetros por segundo en esas cuatro noches. Además, los cambios de longitud de onda de las líneas no identificadas ni siquiera eran coherentes: algunas de las trazas se movían hacia longitudes de onda mayores, en tanto que otras lo hacían simultáneamente hacia longitudes de onda menores.

Al llegar a este punto quedó claro que el comportamiento espectral de SS 433 era bastante más extraño de lo que las observaciones anteriores habían indicado. Con mis colegas de la Universidad de California en Los Angeles, inicié un programa para obtener al menos una breve observación espectral de SS 433 durante todas las noches posibles. Pese a la aguda escasez de tiempo de observación en los grandes telescopios de investigación, esto pudo hacerse gracias al brillo, alto en proporción, del objeto. Por ejemplo, la sensible instrumentación espectroscópica, controlada por ordenador, del reflector Lick de tres metros está proyectada para observar galaxias y quasars muy distantes, objetos de brillo cientos de veces más débil que el de SS 433. Ese tipo de telescopios permite conseguir un espectrograma de buena calidad de nuestra estrella en unos diez minutos y, por tanto, podíamos obtener frecuentes espectrogramas sin alterar otros programas de observación planificados con anterioridad.

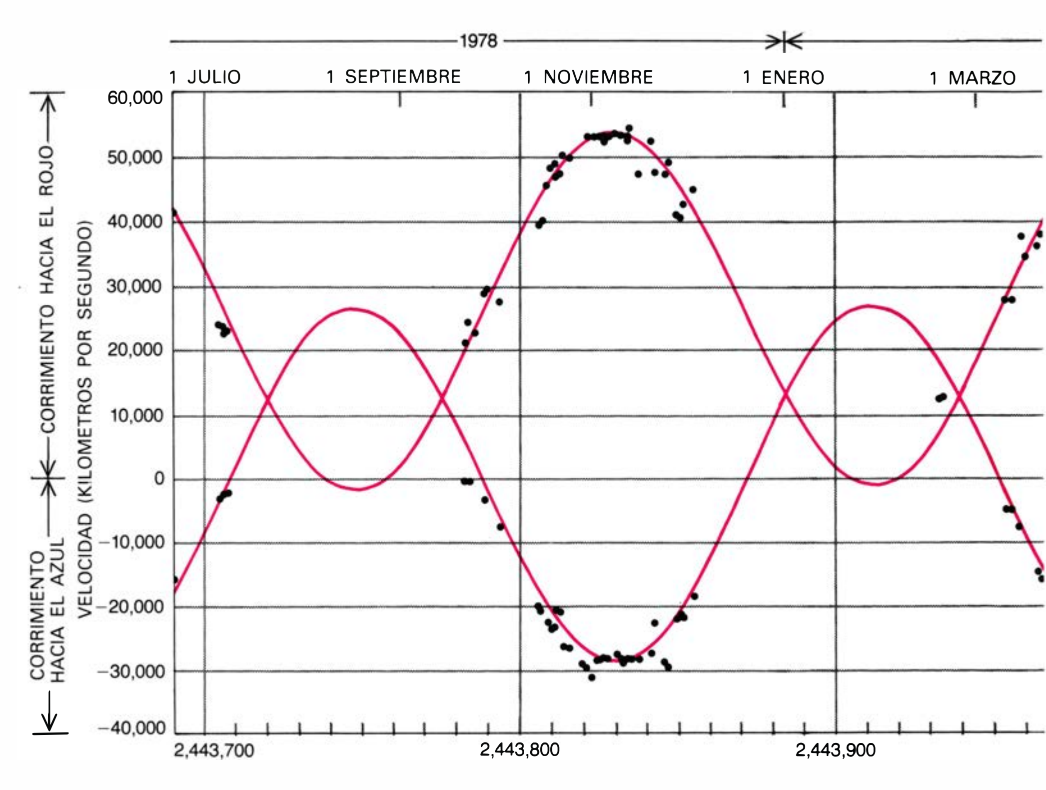
Participaron en las observaciones astrónomos de las cuatro Universidades de California –Los Angeles, Berkeley, Santa Cruz y San Diego– en las que se está llevando a cabo investiga-

ción en astronomía óptica. Más de una docena de astrónomos, cuyos intereses principales variaban desde galaxias distantes hasta estrellas normales cercanas, cedieron generosamente parte de su escaso tiempo de observación para ayudar a vigilar el objeto. Son merecedores de especial mención tres observadores que obtuvieron tantos espectrogramas como yo mismo: Steven A. Grandi y Holland C. Ford, de la UCLA, y Remington P. S. Stone, de la plantilla investigadora de Lick.

Nuestro consorcio tenía poco tiempo que perder, porque, desde diciembre hasta febrero de cada año, la visual de SS 433 está demasiado próxima al Sol para las observaciones nocturnas. Hacia el final del período de observación de 1978, habíamos visto las líneas espectrales "móviles", como las llamábamos, viajar a través de una sorprendente variedad de longitudes de onda. Por ejemplo, la línea de emisión más roja varió su longitud de onda en 700 angstrom en 30 días, lo que, si se interpretase como un corrimiento Doppler gradualmente creciente, implicaría un aumento constante de velocidad desde 20.000 a 50.000 kilómetros por segundo.

En nuestros informes iniciales sobre los datos para la comunidad científica, tratamos de mostrarnos escépticos acerca de la interpretación de las líneas móviles en términos de corrimiento Doppler por diversas razones que en aquella época parecían válidas. Las objeciones más inmediatas eran las que ya he mencionado. Tanto las velocidades necesarias como las variaciones de las velocidades eran unas 100 veces mayores que los valores comparables hallados en cualquier otro objeto estelar. Había también dificultades más sutiles en la explicación del fenómeno como un corrimiento Doppler. Algunas de las líneas móviles se corrían hacia mayores longitudes de onda y otras lo hacían hacia menores, desafiando la sencilla interpretación de una nube de gas que se aproximaba o bien se alejaba.

Además, la identificación de las líneas móviles de emisión como líneas de hidrógeno corridas por efecto Doppler implicaría la existencia de gas a una temperatura bastante modesta, de unos 20.000 grados Kelvin o menos; a temperaturas más altas, la mayor parte del hidrógeno estaría ionizado, y las líneas espectrales debidas a las transiciones de un electrón ligado a un núcleo atómico no podrían aparecer. Sin embargo, casi cualquier mecanismo imaginable que pudiera acelerar el gas a las enormes velocidades implicadas (una

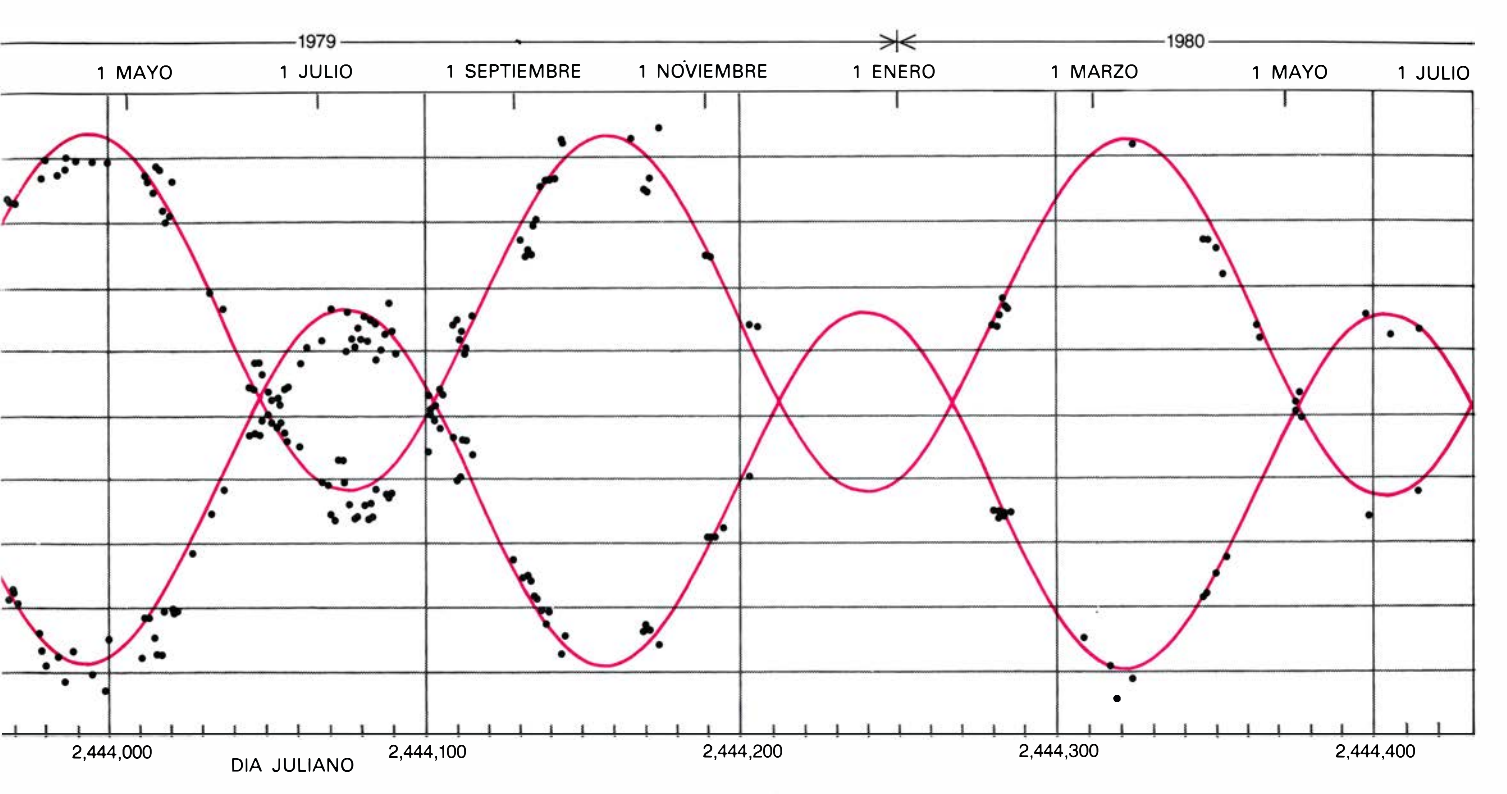


DOS AÑOS DE OBSERVACIONES, resumidos aquí. Se han representado los valores de los corrimientos hacia el rojo y hacia el azul de las líneas de emisión en términos de la velocidad equivalente del gas expulsado. Los grandes intervalos entre los datos corresponden a los tiempos en los que la visual del

fracción substancial de la velocidad de la luz) calentaría el gas hasta temperaturas muy superiores. Otra manera de formular esta dificultad es la siguiente: si las energías térmica y cinética del núcleo de hidrógeno en el gas emisor son aproximadamente iguales, como es caso frecuente en cierto número de sistemas físicos, las velocidades observadas implican temperaturas de más de 30.000 millones de grados Kelvin, mucho más altas que las deducidas para el gas observado en SS 433.

Nuestras preocupaciones no arredraron a un enjambre de imaginativos teóricos. Andrew Fabian y Martin Rees, de Cambridge, indicaron que podría admitirse la existencia de un gas que poseyera corrimientos Doppler en ambos sentidos si un objeto central estuviera emitiendo dos chorros de gas en direcciones aproximadamente opuestas. Parte del gas estaría entonces acercándose al observador y, otra, alejándose. Fabian y Rees hicieron notar la existencia de estructuras con dos lóbulos, que se habían observado en una escala enormemente mayor en radiogalaxias, en las que la radioemisión se encuentra con frecuencia confinada a dos chorros opuestos. Independientemente, Mordechai Milgrom, del Instituto Weizmann de Ciencias de Israel, había propuesto un esquema semejante que iba más lejos, hasta el punto de hacer una predicción que después demostró ser espectacularmente correcta. Basándose sólo en un pequeño número de nuestros datos, Milgrom especuló que los movimientos de las líneas podrían ser periódicos, con un tiempo de repetición de unos pocos meses.

Mientras tanto, nosotros continuamos dándole vueltas a la idea básica en todo su alcance. ¿Eran las líneas móviles no identificadas verdaderamente líneas de hidrógeno corridas por efecto Doppler? La respuesta se reveló con transparencia en marzo de 1979, cuando SS 433 estaba de nuevo lo suficientemente lejos del Sol para la observación espectroscópica. Nuestro primer espectrograma del nuevo período de observación mostró claramente todas las principales líneas de emisión, triplicadas: una de ellas a la longitud de onda de laboratorio (sin corrimiento), otra a una longitud de onda corrida hacia el extremo rojo del espectro (hacia mayores longitudes de onda) y una tercera corrida hacia el extremo azul (hacia longitudes de onda más cortas). Tanto las líneas de emisón del hidrógeno como las del helio, aproximadamente una media docena de trazas en total, exhibían esta peculiar configuración triple. Además, cada una de las líneas corridas hacia el rojo implicaba independientemente una velocidad de recesión idéntica, de unos



objeto estaba demasiado próxima al Sol para que se pudieran realizar observaciones nocturnas; los intervalos más cortos son atribuibles a la proximidad de la Luna, que hace la observación difícil, o a falta de tiempo para observación en un telescopio adecuado. Las curvas muestran el comportamiento predicho, basado en la hipótesis de que el gas responsable de las líneas afectadas por corrimiento Doppler está en dos chorros opuestos.

27.000 kilómetros por segundo, y cada una de las líneas corridas hacia el azul implicaba, de manera análoga, una velocidad de acercamiento idéntica, de unos 6000 kilómetros por segundo. Este múltiple conjunto de coincidencias sólo se podía explicar mediante el corrimiento Doppler. Por la misma época y sin haber mediado contacto alguno, un grupo de astrónomos de la Universidad de Arizona, dirigido por James W. Liebert, había llegado a la misma conclusión.

Por curiosa coincidencia, la dificultad en la interpretación de las líneas móviles de SS 433 era una sorprendente repetición del repentino descubrimiento del significado de los espectros de quasars por Maarten Schmidt, veinte años atrás. Los enormes corrimientos hacia el rojo de los quasars, causados por la expansión del universo, también cambiaban la posición de las líneas de emisión respecto a sus longitudes de onda normales en una cuantía tan grande que era difícil identificar los átomos que las producían, a pesar de que las líneas resultaron pertenecer a los elementos más abundantes y familiares del universo. Aunque todos conocíamos la historia del descubrimiento de los quasars, SS 433 presentó un problema inicial de interpretación a causa de dos complicaciones sin precedente: en primer lugar, tan enormes corrimientos hacia el rojo nunca se habían asociado a estrellas dentro de nuestra galaxia, y en segundo, el objeto mostraba evidentemente corrimientos hacia el azul así como corrimientos hacia el rojo.

Puesto que la longitud de onda de las líneas corridas por efecto Doppler cambia rápidamente con el tiempo, síguese que la velocidad, tanto del gas que se acerca como la del que se aleja, varía y cabe entonces preguntar: ¿Presentan esos cambios alguna regularidad? A lo largo de nuestros dos años de observación, hemos acumulado unas 150 medidas separadas de líneas de emisión, tanto corridas hacia el rojo como hacia el azul; la gráfica de los valores de la velocidad necesaria en función del tiempo revela varias características fascinantes del extraño comportamiento de las líneas móviles [*véase la ilustración de estas dos páginas*]. Se advierte inmediatamente que las velocidades son enormes: el gas corrido hacia el rojo (alejándose) alcanza valores de hasta 50.000 kilómetros por segundo (16 por ciento de la velocidad de la luz) en varias ocasiones distintas, y el gas corrido hacia el azul (acercándose) alcanza velocidades de hasta 30.000 kilómetros por segundo. Como las velocidades de las estrellas en nuestra galaxia (tanto acercándose como alejándose) nunca exceden de unos pocos cientos de kilómetros por segundo, y como todos los objetos extragalácticos más allá de la inmediata vecindad de nuestra galaxia sólo presentan corrimientos hacia el rojo atribuibles a la expansión del universo, SS 433 se distingue por exhibir el mayor corrimiento hacia el azul (en un factor de 100) entre todos los objetos celestes conocidos, sean galácticos o extragalácticos.

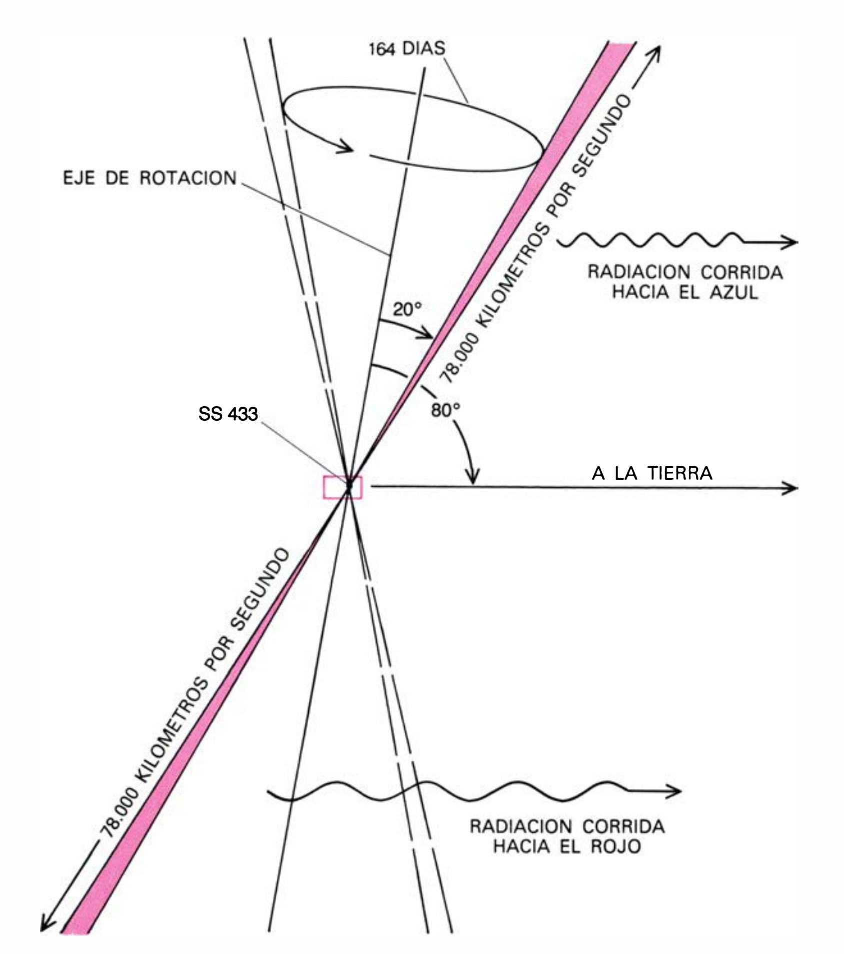
Después de un análisis de la configuración de las variaciones de velocidad, pudimos ordenar más piezas del rompecabezas. Los volúmenes de gas radiante que se acercan y se alejan no coinciden en el espacio; si lo hicieran, se separarían rápidamente al estar esas enormes velocidades dirigidas en sentidos opuestos. Sin embargo, y pese a su separación física, las dos nubes de gas están decididamente relacionadas entre sí; las variaciones en los sistemas corridos hacia el rojo y hacia el azul alcanzan sus valores extremos de velocidad en tiempos idénticos. Además, el valor medio de las dos velocidades para una noche cualquiera, aunque muy grande, resulta casi constante: unos 12.000 kilómetros por segundo.

Podríamos entender la constancia del promedio de las dos velocidades deducidas del corrimiento Doppler, a pesar de las fuertes variaciones de sus valores individuales en una escala de tiempo del orden de días, si un objeto central fuese el responsable de la eyec-

ción de ambas nubes radiantes; la velocidad media sería entonces la de la estrella central. Pero de nuevo nos encontramos con el problema de que este valor excedía por amplio margen la velocidad de escape de nuestra galaxia; el objeto abandonaría la galaxia para siempre en una pequeña fracción de la edad de las estrellas en la vecindad del Sol (al menos 10.000 millones de años). ¿Era razonable creer que uno había tenido la suerte de vivir y trabajar en astronomía durante ese "breve" intervalo? Había una explicación mucho más verosímil.

No he mencionado todavía la que es quizá la característica más sorprendente de la variación de las velocidades de la materia asociada con SS 433. A pesar de las grandes y continuas variaciones de la velocidad, y de los considerables huecos entre las observaciones, del examen de la ilustración de las dos páginas precedentes se puede deducir que tanto el gas corrido hacia el rojo como el corrido hacia el azul vuelven repetidamente a los mismos valores de la velocidad aproximadamente una vez cada seis meses; en otras palabras, las variaciones de la velocidad son periódicas. Un simple análisis de los datos actuales muestra que el período exacto es de 164 días, con una incertidumbre de aproximadamente medio día. El ritmo de la variación se aprecia con toda nitidez. Dos veces cada 164 días, los dos sistemas emisores se reúnen a velocidad idéntica y después se intercambian, esto es, el gas corrido hacia el rojo se convierte en corrido hacia el azul, y viceversa. Desde el punto de vista observacional, estos sucesos se aprecian en el espectro como una reunión gradual de las líneas móviles, que pasan luego una a través de la otra, para separarse en direcciones opuestas. Hemos observado este "cruce" en varias ocasiones, aunque no hemos tenido la suerte de conseguir un espectrograma durante el intervalo preciso (breve, a buen seguro) en el que las líneas móviles están exactamente superpuestas.

Hay un precedente familiar de este tipo de variabilidad espectral periódica, si bien ocurre en escala muy reducida. En ciertos sistemas de estrellas binarias, si el brillo de ambas puede equipararse, podemos observar a veces dos conjuntos independientes de líneas espectrales. Como en cualquier instante una estrella tiene alguna componente de la velocidad hacia el sistema solar y la otra posee alguna componente alejándose de él, los corrimientos Doppler de la misma línea espectral de cada estrella difieren ligeramente; de aquí que las líneas aparezcan como un doblete resoluble a longitudes de onda un poco diferentes. A medida que las dos estrellas giran en torno a su centro de masa común, la componente del movimiento de cada estrella hacia la Tierra o alejándose de ella cambia suave y periódicamente; en consecuencia, las longitudes de onda de las dos líneas espectrales experimentan un cambio periódico entre dos longitudes de onda límites. En los sistemas de estrellas binarias, sin embargo, la amplitud del cambio de la velocidad nunca excede de unos pocos cientos de kilómetros por segundo, mientras que en SS 433 la amplitud es más de 100 veces mayor.



GEOMETRIA DEL MODELO DE CHORROS GIRATORIOS de SS 433, esquematizada. El eje de rotación del objeto está inclinado respecto a la visual del sistema solar un ángulo de unos 80 grados. Los chorros mismos están a su vez inclinados unos 20 grados respecto al eje de rotación. El período de rotación es de 164 días. La velocidad del chorro resulta en este modelo ser de unos 78.000 kilómetros por segundo. La componente de esta velocidad proyectada a lo largo de la visual varía periódicamente al girar el eje de los chorros; de aquí que las líneas de emisión afectadas por el efecto Doppler en el espectro del objeto fluctúen entre posiciones de máximo y mínimo alejamiento, de acuerdo con la inclinación de los chorros respecto a la visual. En esta posición, el corrimiento hacia el rojo y hacia el azul serían máximos; los mínimos corrimientos se apreciarían cuando los chorros se hallaran perpendiculares a la visual.

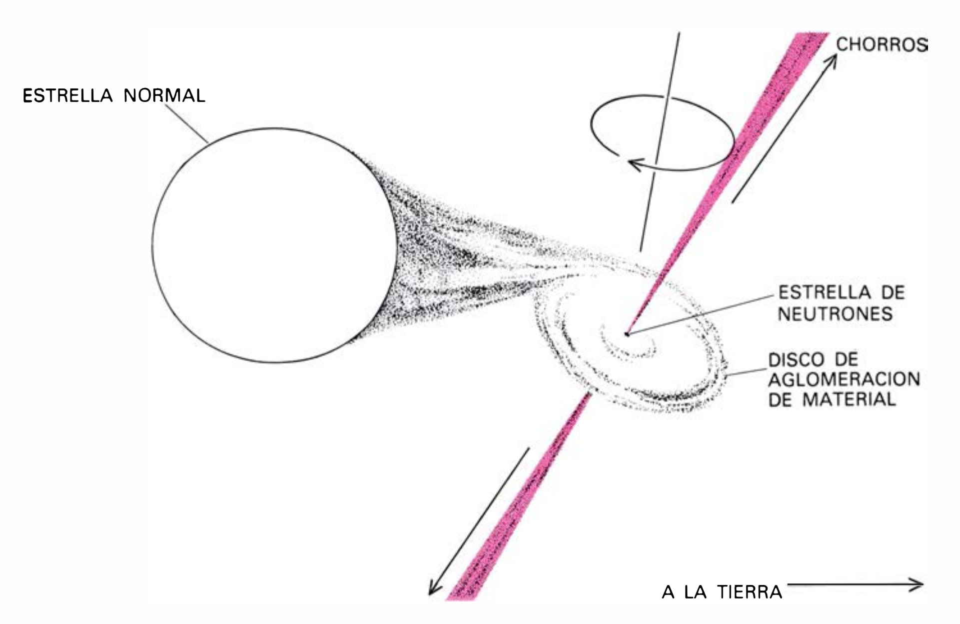
¿Podría SS 433 representar una versión, muy ampliada, de este fenómeno, constando de sólo dos objetos que describen órbitas uno en torno del otro? Si este fuera el caso, se deduciría una fascinante consecuencia. Un corolario de la sencillez de la fuerza gravitatoria es que todos los cuerpos en órbita, con independencia de su naturaleza, obedecen una relación básica entre el período orbital, la velocidad orbital y la masa total del sistema. Por tanto, si se supone que las espectaculares variaciones de longitud de onda en SS 433 se deben a movimiento orbital, basta con la amplitud total de la variación y el período de la variación de velocidad para calcular directamente la cantidad total de materia en SS 433. ¡El resultado es una masa equivalente a mil millones de veces la del Sol! Si pensamos que las estrellas de mayor masa conocida no llegan a las 100 masas solares, este resultado nos parecerá asombroso. En efecto, la masa así deducida para SS 433 viene a ser del 1 por cien de la masa total de los 100.000 millones de estrellas de la totalidad de la galaxia.

¿Cabe suponer que un objeto de tan extraordinaria masa haya escapado a la detección hasta ahora? La respuesta es negativa. Hay varios argumentos distintos que demuestran que las extrañas variaciones de las líneas espectrales no

pueden deberse a movimiento orbital. Por ejemplo, los supuestos parámetros orbitales son de un orden muy elevado, hasta el punto de que la órbita habría de tener un diámetro de tan gran magnitud que la luz tardaría dos semanas en el viaje entre los objetos en órbita. A pesar de todo, los máximos y mínimos en la gráfica de los corrimientos hacia el rojo y hacia el azul del sistema permanecen sincronizados entre sí con una discrepancia inferior a uno o dos días. De acuerdo con la teoría especial de la relatividad, no es posible propagar información alguna, independientemente del método de transmisión, con velocidad superior a la de la luz. No había pues manera de que los objetos permaneciesen sincronizados; no podrían, si se nos permite la expresión, "decirse" uno al otro dónde se encuentran.

Una segunda indicación que arguye convincentemente en contra del movimiento orbital es el sistema estacionario de emisión en el espectro de SS 433, es decir, el conjunto de líneas espectrales que están constantemente muy cerca de su longitud de onda medida en el laboratorio. Si el gas que emite esas líneas estuviera en la vecindad de un objeto con una masa de mil millones de masas solares, respondería en seguida a esa enorme atracción gravitatoria y caería hacia la masa grande. Sin embargo, las velocidades indicadas por esas líneas son muy pequeñas.

Por todas esas razones, unidas a ciertos argumentos más sutiles, no es factible invocar la presencia de dos cuerpos, cada uno describiendo una órbita en torno al otro, para explicar los rápidos cambios en longitud de onda en el espectro de SS 433. En vez de seguir ese camino, la mayoría de los investigadores interesados en el problema han retornado a modificaciones del concepto según el cual un objeto central emite dos chorros de materia, uno de ellos dirigido aproximadamente hacia la Tierra, dando así lugar a las líneas de emisión corridas hacia el azul, y otro chorro dirigido casi en sentido opuesto, que da lugar a la radiación corrida hacia el rojo. Si se postula además que la línea imaginaria que une los dos chorros gira con una velocidad tal que la revolución completa dura 164 días, se explican también las modulaciones periódicas de los valores observados de la velocidad. Esto se debe a que, en cada punto del ciclo de 164 días, el ángulo de uno y otro chorro respecto a la visual de la Tierra varía. Cuando los chorros están en la posición más próxima a la visual terrestre (apuntando hacia ella, o al revés), se observan las máximas velocidades de aproximación o alejamiento. Por otra parte, cuando los chorros están dirigidos perpendicularmente a la visual terrestre, el gas no se acerca ni se aleja de la Tierra, y es de esperar que las velocidades sean mínimas.



HIPOTETICOS OBJETOS CENTRALES POSTULADOS por el autor y sus colegas para ayudar a explicar las características observadas del espectro de SS 433, según el modelo de chorros giratorios; este diagrama, que los representa, puede considerarse como una vista ampliada de la región situada dentro del pequeño encuadre en color en la ilustración de la página opuesta. El objeto responsable de la expulsión de los chorros se estima es parte de un sistema binario de estrellas, que consta de una estrella comparativamente normal (*izquierda*), no muy distinta del Sol, ligada por una órbita próxima a una compacta estrella de neutrones (*derecha*), la cual se halla en el proceso de arrancar materia de la compañera normal por efecto de su intenso campo gravitatorio. (Se supone que ambas estrellas poseen aproximadamente la masa del Sol.) El gas procedente de la estrella normal forma un disco giratorio de acumulación alrededor de la estrella de neutrones, y los dos chorros salen expulsados en direcciones opuestas desde las caras de este disco. La precesión del plano del disco giratorio de acumulación alrededor de la estrella de neutrones es, presumiblemente, lo que hace girar el eje de los chorros.

Esa idea admite una fácil cuantificación mediante un sencillo sistema de ecuaciones. Por sorprendente que resulte, sólo hay cinco incógnitas en las ecuaciones; en otras palabras, bastan cinco variables características para describir todo el comportamiento del objeto. Para empezar, hay dos ángulos geométricos desconocidos. Uno es la inclinación del eje de rotación del objeto respecto a la visual; así, un observador extraterrestre capaz de percibir la rotación de la Tierra no estaría obligado a situarse necesariamente en la vertical del Ecuador; percibiría la rotación desde cualquier latitud. No sabemos la "latitud" desde la cual estamos observando SS 433. La segunda incógnita es la inclinación del eje de los chorros respecto al eje de rotación. Nuestro ángulo respecto a los chorros no tiene por qué ser este primer ángulo, de la misma manera que los ejes de los polos geográficos y magnéticos en la Tierra están ligeramente inclinados uno respecto al otro. La velocidad con que se expulsan los chorros constituye una tercera incógnita; por razón de sencillez, es conveniente suponer que los dos chorros son expulsados en direcciones opuestas, aunque con la misma velocidad. (Si esta hipótesis es errónea, las ecuaciones no pueden resolverse.) Finalmente, los valores precisos del período y la fase del ciclo de 164 días de SS 433 son ambos incógnitas que deben determinarse a partir de los datos.

Nuestras observaciones espectroscópicas pueden ajustarse a este modelo teórico para resolver si pueden hallarse valores de las cinco variables que estén de acuerdo con los datos. Resulta que este sencillo concepto, de manera quizá sorprendente, concuerda muy bien con los datos. Los dos ángulos resultan ser respectivamente 80 y 20 grados, y la velocidad del chorro exigida por las observaciones es de 78.000 kilómetros por segundo, o 26 por ciento de la velocidad de la luz. La razón por la cual nunca observamos directamente un corrimiento hacia el rojo o hacia el azul, de un valor tan alto como éste, se debe a que los dos ángulos geométricos no son rectos. Con ello, los chorros no apuntan directamente hacia, o alejándose de, la Tierra. (En realidad, sería sospechoso que así lo hicieran.) La máxima componente de la velocidad del gas moviéndose hacia la Tierra o

alejándose de ella no puede, por tanto, alcanzar el valor real de la velocidad deducido para los chorros. Un observador que se hallara en un lugar a lo largo de los chorros sí observaría estas velocidades máxima y mínima.

Para la materia que se está moviendo a una fracción importante de la velocidad de la luz, los efectos predichos por la teoría especial de la relatividad cobran también importancia. De hecho, uno de tales efectos explica uno de los aspectos más intrigantes de SS 433: la velocidad media, constante y muy grande (12.000 kilómetros por segundo), de los dos chorros. Aunque el efecto pertinente tiene distintos nombres técnicos, como corrimiento Doppler transversal o de segundo orden, es frecuente designarlo por dilatación del tiempo. Un observador imaginario que observase a alguien moviéndose con gran velocidad y transportando un reloj, percibiría que el reloj se atrasaba y tanto más cuanto mayor fuera la velocidad. La persona que llevara el reloj, por el contrario, apreciaría que el reloj marchaba perfectamente. En realidad, como todo movimiento es relativo, podría acusar al observador "estacionario" de percepción incorrecta del tiempo.

¿Qué tiene esto que ver con SS 433? Cada átomo del gas expulsado viene a ser como un reloj, pues debe vigilar la frecuencia (o, en forma equivalente, la longitud de onda) y emitir luz a la longitud de onda apropiada siempre que sus electrones son desexcitados. Ahora bien, la dilatación del tiempo siempre retrasa el reloj y la frecuencia y la longitud de onda son inversamente proporcionales entre sí; por tanto, el decrecimiento de frecuencia se reflejará en un aumento de la longitud de onda o, en otras palabras, en un corrimiento hacia el rojo. La cuantía del corrimiento hacia el rojo, por la dilatación del tiempo, depende sólo de la velocidad del reloj; y es fácil calcular que, a la velocidad de aproximadamente un cuarto de la de la luz, el efecto de la dilatación del tiempo es de un 4 por ciento.

Quizá nos parezca ello una pequeña discrepancia, pero recuérdese que un corrimiento hacia el rojo del 4 por ciento de la velocidad de la luz (300.000 kilómetros por segundo) es 12.000 kilómetros por segundo. Este es, desde luego, exactamente el valor medio observado de los dos haces de SS 433. El corrimiento hacia el rojo debido a la dilatación del tiempo siempre aparece en los chorros, con independencia de su ángulo respecto a la visual. Por tanto, el corrimiento Doppler variable, con un período de 164 días, está superpuesto sobre (sumándose a y restándose de) el corrimiento hacia el rojo causado por dilatación del tiempo. Esta interpretación no sólo explica con rigor por qué la velocidad media observada de los dos chorros es de 12.000 kilómetros por segundo, sino que da cuenta también de por qué las líneas de emisión asociadas a los dos chorros se reúnen dos veces cada 164 días en ese valor grande y no en el correspondiente a velocidad nula. Los cruces tienen lugar cuando los chorros están situados perpendicularmente a la visual. Aun cuando en esos instantes no hay acercamiento o alejamiento del gas emisor, y no existe por tanto corrimiento hacia el rojo o hacia el azul distinto del de la dilatación del tiempo, este omnipresente efecto dota a ambos chorros de un corrimiento hacia el rojo equivalente a 12.000 kilómetros por segundo.

¿Qué es, pues, el misterioso objeto central que emite los chorros? Aunque no hay contestación sencilla, tenemos una indicación importante: la velocidad observada, del 26 por ciento de la velocidad de la luz. ¿Por qué es este valor y no cualquier otro? La respuesta puede estribar en que esa velocidad es muy próxima a la velocidad de escape de la materia desde la superficie de una estrella de neutrones. Los cálculos teóricos muestran que hay un margen limitado de variables dentro del cual pueden esas extrañas estrellas soportar su propia masa y así ser estables. Una estrella de neutrones con una masa igual a la del Sol tendría un radio de unos 10 kilómetros tan sólo. La velocidad requerida para escapar de la superficie de una estrella de neutrones resulta ser similar a la observada en los chorros de SS 433. Tal vez se trate de una coincidencia; y, si no lo es, sugiere la posibilidad de un mecanismo de expulsión autorregulado. El gas puede acelerarse hasta la velocidad necesaria para expelerlo permanentemente; entonces, el mecanismo acelerador, que ha cumplido ya su función, no necesita trabajar con más intensidad. Podríamos invertir el argumento. Si la estrella no posee esa compacidad ¿por qué habría de generar y mantener las enormes velocidades de expulsión observadas, cuando dotando al material de una velocidad mucho más baja podría librarse de él para siempre?

En el caso de que SS 433 aloje realmente una estrella de neutrones, el reloj de 164 días (es decir, el mecanismo que hace girar el eje del chorro) es, con probabilidad, distinto de la simple rotación de la estrella cada 164 días. Y ello es así porque el tamaño, muy pequeño, de la estrella implica que a esa baja velocidad de rotación el objeto no sea muy rígido. (Usando una expresión técnica diríamos que su momento de inercia no es muy grande.) Por otra parte, la materia expulsada se lleva consigo una tremenda cantidad de energía a una velocidad fantástica. El retroceso resultante impartido a la estrella destruiría de inmediato el comportamiento periódico, a menos que las dos masas de materia eyectada tuvieran exactamente la misma velocidad y se hallaran alineadas, de modo que sus efectos sobre la estrella se anularan entre sí. Parece más probable que el período de 164 días sea, en cambio, un efecto de precesión, esto es, una lenta variación del eje de rotación. El verdadero período de rotación de la estrella podría ser corto, como ocurre en los radiopulsares, donde los períodos son del orden de algunos segundos o menos. Como ejemplo análogo podríamos pensar en un giroscopio de juguete, cuya rueda puede girar muy deprisa mientras que el conjunto describe una circunferencia poco a poco.

¿Cuál es la fuente de la materia expulsada? Las líneas de emisión son de hidrógeno y helio, elementos hallados típicamente en objetos mucho menos evolucionados que una estrella de neutrones; por consiguiente, la materia probablemente no provendrá de tal estrella. El medio interestelar consta en gran medida, de hidrógeno y helio, pero es demasiado tenue para suministrar suficiente gas. Ello nos induce a sospechar la presencia de una segunda estrella menos evolucionada.

Especulando en esa línea de razonamiento, David Crampton, Anne P. Cowley y John B. Hutchings, del Observatorio Astrofísico del Dominio en Canadá, se propusieron obtener una serie de espectrogramas de SS 433 que pudieran ser sensibles a tal estrella compañera. Pronto descubrieron que las líneas "estacionarias" de emisión, esto es, líneas de emisión del hidrógeno y helio a sus longitudes de onda de laboratorio, mostraban también un corrimiento cíclico en su longitud de onda. La amplitud del corrimiento era muy pequeña, correspondía a 70 kilómetros por segundo (alrededor del 0,1 por ciento de la amplitud del corrimiento de las líneas móviles), y por tanto no podía detectarse con el equipo que mis colegas y yo usábamos en nuestras observaciones. El período de la variación menor era de 13 días. De nuevo, el período observado y la

amplitud de la variación de la velocidad conducía a estimas de las masas de las dos estrellas. Se llegaba a una solución coherente, aunque no única, si se suponía que las dos estrellas poseían una masa comparable a la del Sol. Los parámetros de la solución orbital sugerían que la compañera de la estrella de neutrones era probablemente una estrella normal, de características no muy diferentes de la del Sol. Por desgracia, tal objeto sería demasiado débil frente al fondo de radiación de los chorros para observarse directamente.

Para que no dé la impresión de que se han desvelado todos los misterios de SS 433, voy a pasar revista al enjambre de intrigantes problemas pendientes de solución. La mayoría de ellos se centran en torno a las condiciones físicas del material expulsado. Cabe calcular algunas de las variables del gas emisor; para hacerlo, sin embargo, es preciso formular alguna hipótesis acerca de la distancia de SS 433, de suerte que la intensidad observada de los chorros de radiación pueda usarse para calcular la luminosidad intrínseca. Al igual que la mayoría de las estimas astronómicas, ésta dista mucho de tener total garantía de certeza. Sobre la base de argumentos tales como la intensidad de las absorciones espectrales debidas al gas interestelar interpuesto, he estimado la distancia en unos 4000 parsecs. (Un parsec vale 3258 años-luz.) El diámetro de nuestra galaxia mide unos 30.000 parsecs, de modo que SS 433, aunque no sea un vecino inmediato del Sol, tampoco se halla muy alejado.

Podemos calcular ya la luminosidad de los chorros. Las pocas líneas de emisión móviles prueban tener unas 10 veces más energía que el Sol radia a todas las longitudes de onda. La longitud de la materia visiblemente radiante es de unos 10.000 millones de kilómetros, esto es, unas 100 veces la distancia de la Tierra al Sol, o dos veces el tamaño de todo el sistema solar. Pero la variable más desconcertante es la cantidad de energía cinética que, según los cálculos, se necesita para acelerar la considerable masa de materia a la extraordinaria velocidad observada. Aunque la cifra depende algo de hipótesis dudosas, es del orden de 10^{39} erg por segundo, o bien un millón de veces la energía radiada a todas las longitudes de onda cada segundo por el Sol. ¿Cuál es la fuente de esta fantástica producción de energía? Nada sabemos de ella con seguridad.

Cuanto más concretas son las preguntas que hacemos acerca de los chorros, tanto más se revela nuestra ignorancia. ¿Qué mecanismo canaliza esta energía para acelerar el gas, dando precisamente una velocidad constante y con controlada exactitud? ¿Por qué está el gas tan frío en comparación con las temperaturas que se esperarían a esas velocidades? ¿Qué proceso colima y dirige la materia para constituir un chorro? Esta última cuestión incordia particularmente. Disponemos ya de una medición observacional de la anchura de los chorros, basada en la anchura de las líneas espectrales móviles. Si los chorros fueran anchos, en cualquier instante dado una tal corriente presentaría una variedad de ángulos, así como velocidades y corrimientos Doppler diferentes. Sería de esperar entonces que las líneas móviles fuesen anchas, cubriendo este intervalo de velocidades Doppler. En realidad observamos lo contrario: líneas móviles muy estrechas en comparación con la enorme velocidad de los chorros. De lo que se sigue que los chorros tengan una anchura inferior a algunos grados, casi como dos agudos estiletes.

Por último, la pregunta quizá más intrigante: ¿Dónde están los otros objetos semejantes a SS 433? ¿Por qué observamos sólo uno de tales objetos en una galaxia de 100.000 millones de estrellas? Podríamos responder tal vez que la duración de la vida de este extraño suceso es muy corta según los patrones astronómicos, quizás del orden de 10.000 años. Muchas estrellas pueden pasar por esta fase, pero sólo durante un instante astrónomico. Es posible que sólo haya un objeto de esos en actividad en cada momento.

Sería aventurado especular sobre la importancia general de SS 433 para la astrofísica en los próximos cinco o diez años. No niego el interés que reviste tratar de adivinarlo. La posibilidad más sugestiva es la de que la semejanza de la estructura del doble chorro de SS 433 con el doble lóbulo de radioemisión de las galaxias gigantes y los quasars no obedezca a una pura coincidencia. Sería sumamente estimulante que se escondiera el mismo mecanismo básico bajo ambos fenómenos, y que las respuestas a las preguntas anteriores sobre energética, aceleración y colimación estuvieran relacionadas en ambos casos. Tendríamos entonces el privilegio de disponer de una vista en primer plano de un objeto comparativamente cercano dentro de la galaxia, que podría servir de prototipo para mejorar nuestra comprensión de los violentos sucesos extragalácticos, que se cuentan entre los grandes misterios de la astronomía.

Anticuerpos monoclonales

Las células secretoras de anticuerpos pueden hacerse inmortales fusionándolas con células tumorales y clonando los híbridos resultantes. Cada clon producirá, de forma persistente, grandes cantidades de un anticuerpo, altamente específico

Cesar Milstein

Cuando una sustancia extraña penetra, o se inyecta, en el cuerpo de un vertebrado, uno de los aspectos de la consecuente respuesta inmune es la secreción de anticuerpos por parte de ciertas células plasmáticas. Los anticuerpos son moléculas de inmunoglobulina con sitios de combinación que reconocen la forma de los determinantes situados en la superficie de la sustancia extraña, o antígeno, y se unen a ellos. La combinación del anticuerpo con el antígeno desencadena ciertos procesos capaces de neutralizar y eliminar la sustancia extraña. Aunque sin relación directa con esa su función natural en la respuesta inmune, los anticuerpos han constituido también una herramienta importante para los investigadores, que han aprovechado su especificidad para identificar y marcar determinadas células o moléculas y separarlas luego de una mezcla.

La respuesta del anticuerpo a un antígeno típico es muy heterogénea. En el bazo de un ratón o un hombre hay hasta un millón de estirpes celulares distintas de linfocitos *B,* las células precursoras de las células plasmáticas. Aunque todas ellas derivan de un tronco celular común, cada estirpe adquiere la capacidad de fabricar independientemente un anticuerpo que reconoce a un determinante antigénico distinto. Cuando se inyecta un agente inmunizante en un animal, éste responde elaborando anticuerpos dirigidos contra las diferentes moléculas antigénicas del material inyectado y los distintos determinantes de un solo antígeno, y produce también diferentes anticuerpos que se acoplan, más o menos bien, a un mismo determinante. Es poco menos que imposible separar totalmente los distintos anticuerpos, por cuya razón los antisueros convencionales contienen mezclas, que difieren de un animal a otro.

Cada anticuerpo se elabora en una estirpe distinta de linfocitos y células plasmáticas derivadas. ¿Qué ocurriría si separáramos una célula tal que fabricara un único anticuerpo específico para que proliferara en un cultivo? La progenie, o clon, de esta célula sería una fuente de grandes cantidades de anticuerpos idénticos frente a un único determinante antigénico: se trataría de un anticuerpo monoclonal. Para nuestro infortunio, las células secretoras de anticuerpos no pueden mantenerse en un medio de cultivo.

Hay ciertos tumores malignos del sistema inmune, llamados mielomas, cuyas células, que poseen una gran capacidad de proliferación, producen grandes cantidades de inmunoglobulinas anormales, las denominadas proteínas mielómicas. De suyo, el tumor es un clon inmortal de células descendientes de una única progenitora; las células del mieloma pueden cultivarse indefinidamente y todas las inmunoglobulinas que segregue tendrán una estructura química idéntica. Estas inmunoglobulinas son, en efecto, anticuerpos monoclonales, pero no hay forma de saber contra qué antígeno se dirigen, ni de qué modo pueden inducirse mielomas que produzcan anticuerpos frente a un antígeno específico.

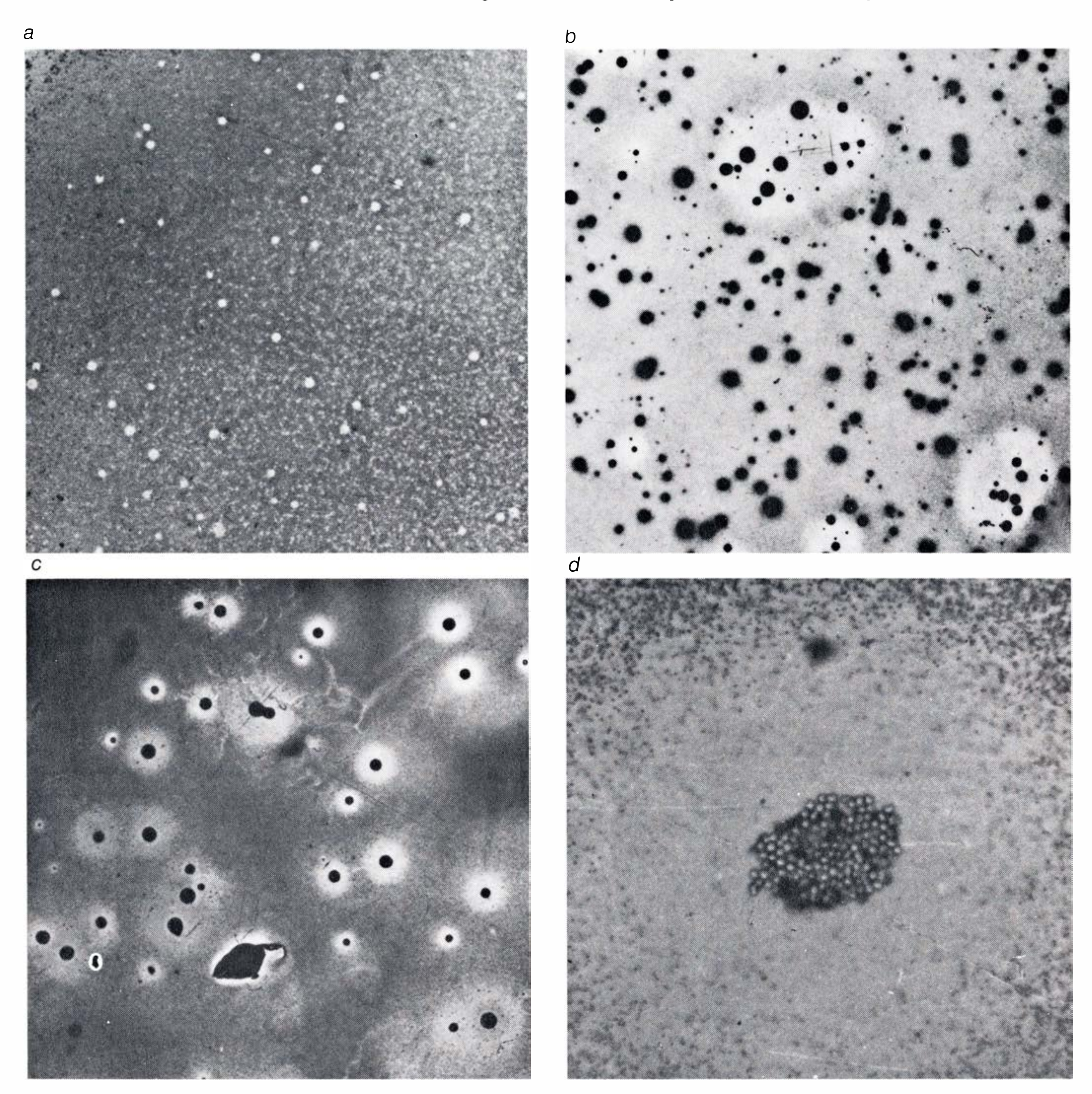
En 1975, mis colegas y yo conseguimos fusionar células de mieloma de ratón con linfocitos de bazo de ratones inmunizados con un determinado antígeno. Las células del mieloma híbrido resultante, o "hibridoma", expresaban tanto la propiedad del linfocito de producir anticuerpos específicos como el carácter inmortal de las células mielómicas. Estas células híbridas pueden manipularse mediante las técnicas aplicables a los cultivos permanentes de células animales. Cada célula híbrida puede clonarse; y cada clon produce grandes cantidades de un anticuerpo específico para un solo determinante antigénico. A su vez, los clones pueden conservarse indefinidamente y en cualquier momento pueden tomarse muestras de ellos para su cultivo o inyección en animales a fin de obtener anticuerpo monoclonal a gran escala. Los anticuerpos altamente específicos producidos por este método general han resultado ser herramientas de notable versatilidad en muchas áreas de la investigación biológica y la clínica médica.

Los mielomas humanos se conocen en medicina desde hace tiempo, pero hasta los primeros años de la década de 1960 los inmunólogos no consiguieron desvelar la naturaleza precisa de las proteínas mielómicas. Por esas fechas, Michael Potter, del National Cancer Institute, indujo mielomas en ratones, mielomas que produjeron grandes cantidades de inmunoglobulinas monoclonales. A pesar de los muchos intentos, no se logró inducir tumores que sintetizaran anticuerpos contra un antígeno inyectado. Leo Sachs, Kenko Horibata, Edwin S. Lennox y Melvin Cohn, del Salk Institute for Biological Studies, consiguieron mantener en cultivo una estirpe celular de mieloma de ratón; mas la estirpe se perdió. Por fin, Horibata y A. W. Harris establecieron varias estirpes, que distribuyeron a otros laboratorios. Mi grupo, del Medical Research Council Laboratory of Molecular Biology, de Cambridge, sometió a estudio intensivo una estirpe celular derivada de uno de los tumores producidos por Potter.

Entonces no nos preocupaban los anticuerpos monoclonales. Estábamos investigando cómo se diversificaban las células somáticas en cultivo y de qué manera las mutaciones modificaban la especificidad de combinación de los anticuerpos; los mielomas de ratón no eran para nosotros más que otra estirpe celular mantenida en cultivo. En 1973, Richard G. Cotton, David S. Secher y el autor lograron, por primera vez, producir mutantes estructurales de una proteína de mieloma de ratón segregada por una estirpe celular cultivada.

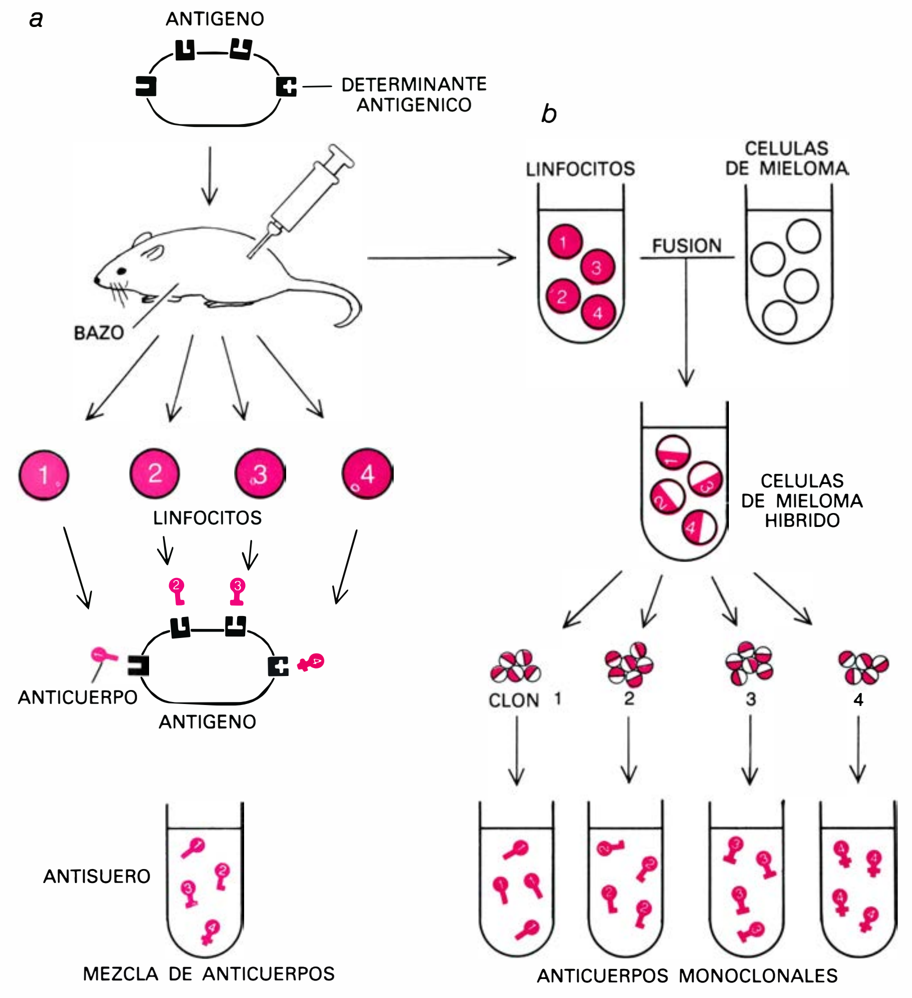
Aquel trabajo e investigaciones paralelas de Matthew D. Scharff, de la Facultad de Medicina Albert Einstein de Nueva York, demostraron la aparición en células cultivadas de mutaciones espontáneas que afectaban a la estructura de las proteínas fabricadas por las células, aportando, además, datos sobre la frecuencia y naturaleza molecular de esas mutaciones. La búsqueda de mutantes fue, no obstante, laboriosa, pues las proteínas fabricadas por las células progenitoras carecían de actividad como anticuerpo; las variaciones de esta actividad hubieran sido la señal más evidente de la aparición de pequeñas diferencias derivadas de las mutaciones. A todas luces, se postulaba una estirpe celular que segregara una inmunoglobulina dotada de capacidad de anticuerpo, y cuya actividad pudiera someterse fácilmente a ensayo. Pero esa estirpe no existía en aquel entonces.

Así las cosas, una circunstancia afortunada nos condujo a la técnica del mieloma híbrido. Aunque trabajábamos sobre mutaciones somáticas, Georges Köhler y yo seguíamos una línea de investigación bastante diferen-

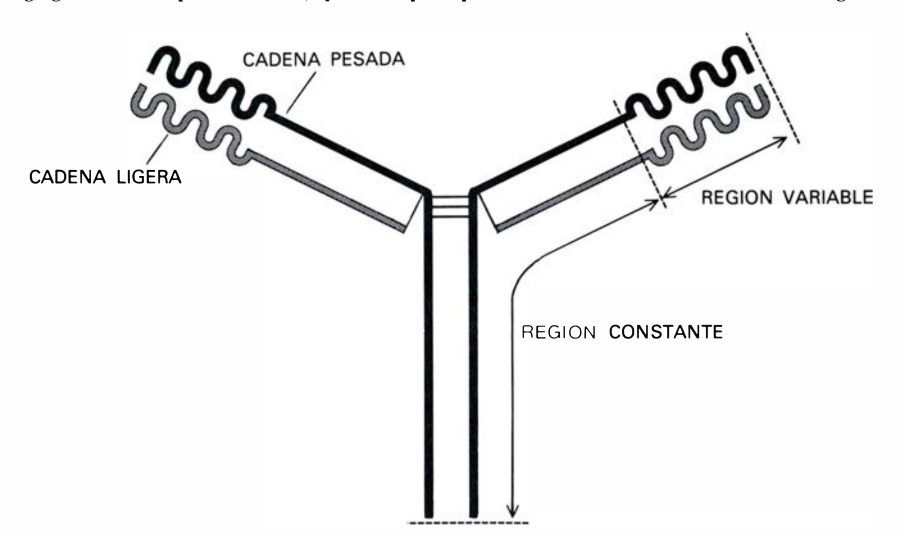


CLONES SECRETORES DE ANTICUERPOS derivados de células híbridas de mieloma. Se detectaron por primera vez mediante un ensayo de anticuerpos contra eritrocitos de oveja. En la prueba convencional (*arriba, a la izquierda*), los eritrocitos y las células productoras de anticuerpos se incuban en agar y se añade el complemento (un complejo proteico del plasma sanguíneo). El anticuerpo se difunde desde las células secretoras y se enlaza a los antígenos de las células sanguíneas más próximas. Se inicia así la reacción del complemento, que provoca la muerte de las células sanguíneas y da lugar a la aparición de placas (*manchas blancas*) alrededor de cada célula secretora.

El autor y George Köhler fusionaron células de ratón, secretoras de anticuerpo frente a células sanguíneas, con células tumorales (mielómicas) y sembraron los híbridos (*arriba, a la derecha*). Aparecieron colonias híbridas (*manchas negras*). Al añadir una capa de eritrocitos de oveja junto con complemento, unas cuantas colonias, las que segregaban anticuerpo específico, dieron lugar a placas (*zonas blancas alrededor de las colonias*). Se tomaron células de una colonia secretora de anticuerpo y se sembraron (*abajo, a la izquierda*); la mayoría de los clones resultantes secretaban anticuerpo, como prueba el área de células muertas que rodea al clon (*abajo derecha*).



RESPUESTA INMUNE que se inicia (*a*) cuando una molécula de antígeno, portadora de varios determinantes antigénicos diferentes, penetra en el cuerpo del animal. La respuesta del sistema inmune implica una proliferación de estirpes de linfocitos *B*. Cada estirpe segrega moléculas de inmunoglobulinas que se acoplan a un único determinante antigénico, o a una parte del mismo. El antisuero convencional contiene una mezcla de estos anticuerpos. Los anticuerpos monoclonales proceden de linfocitos de bazo fusionados con células de mieloma maligno (*b*). Se clonan las células híbridas individuales; cada clon segrega un anticuerpo monoclonal, que se acopla específicamente a un único determinante antigénico.



LOS ANTICUERPOS, mostrados en la ilustración superior con formas estilizadas, pertenecen a la familia de proteínas llamadas inmunoglobulinas. La forma básica de una molécula de inmunoglobulina corresponde a la de la clase G (IgG), población heterogénea de moléculas que tienen en común una estructura en forma de *Y*, compuesta por dos clases de cadenas (una pesada y otra ligera), unidas por un enlace disulfuro. El número y la posición precisa de estos enlaces difieren, y son característicos de las subclases de inmunoglobulinas G. Cada cadena tiene dos regiones. En la región variable, las secuencias de aminoácidos, que difieren de un anticuerpo a otro, dan lugar a sitios combinantes, de formas distintas, que encajan específicamente en los distintos antígenos. La región constante de las cadenas, con idéntica secuencia de aminoácidos en todos los anticuerpos de una subclase dada, tiene otras funciones a su cargo.

te para conocer mejor el control genético de la síntesis de anticuerpos. La síntesis de anticuerpos está regulada por dos conjuntos de genes; uno de ellos codifica la región "variable" de las cadenas ligera y pesada de la molécula de anticuerpo, región responsable de su especificidad; y el otro determina la región constante, de la que dependen funciones realizadoras tales como el enlace del complemento (un complejo de proteínas del plasma sanguíneo implicado en la respuesta inmune), el transporte de la molécula de anticuerpo a través de las membranas y su unión a ellas. Cada linfocito sintetiza un anticuerpo codificado por un único par de genes *V* (de variable) y *C* (de constante), de entre el gran repertorio de tales genes que hay en la célula. Cuando existen diferentes alelos, formas variables de los genes *V* o *C*, en cada uno de los cromosomas de la célula, sólo el alelo de un cromosoma se muestra activo; el otro queda excluido.

En 1973, Cotton y yo llevamos a cabo un experimento para determinar si podía evitarse la exclusión alélica, y cuáles serían sus consecuencias a nivel molecular. Fusionamos dos células de mieloma, una de ratón y otra de rata. El análisis de las células híbridas mostró que segregaban moléculas híbridas, es decir, moléculas que eran combinación de las cadenas sintetizadas por las células progenitoras. Ahora bien, nunca se combinaba una región *V* de un animal con una *C* del otro. Este resultado demostraba que las regiones codificadoras para *V* y *C* estaban situadas en el mismo cromosoma. Sabemos ahora que las secuencias de ADN que codifican para las regiones *V* y *C* están separadas por intrones, segmentos intermedios de ADN. El fragmento completo de ADN se transcribe en un filamento complementario de ARN nuclear. Este ARN madura por la acción de enzimas celulares –el intrón se elimina y se empalman las secuencias correspondientes a las regiones *V* y *C*– resultando el ARN mensajero que será traducido a la cadena proteica de la inmunoglobulina. En otras palabras, el experimento del mieloma híbrido demostró que las secuencias para *V* y *C* se transcribían en una sola cadena de ARN.

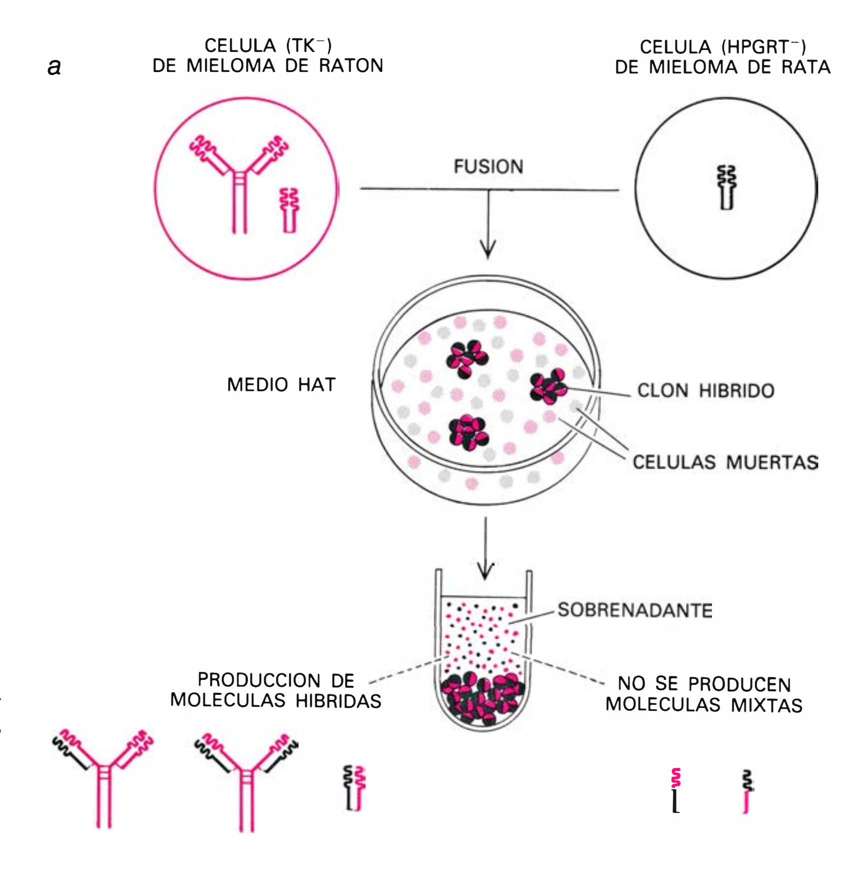
También demostró este experimento que no existía exclusión alélica en los híbridos, pues la información de las dos células progenitoras se expresaba "codominantemente" en las células fu-

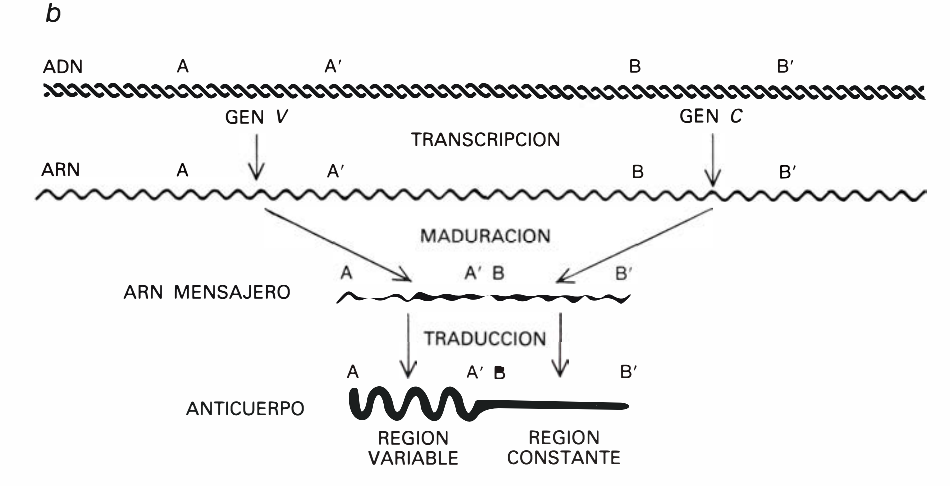
sionadas. Este hallazgo nos sugirió una posible respuesta a la necesidad de una célula productora de anticuerpos para el experimento de mutación. Se nos ocurrió fusionar un linfocito, o célula plasmática, normal con una célula mielómica, inmortalizando de este modo la expresión de la secreción de anticuerpo específico de la célula plasmática. Aplicábamos así la bien conocida técnica de fusión celular a un nuevo propósito: fijar en una estirpe celular permanente una función que sólo se expresa normalmente en una célula "terminal": la célula plasmática derivada de un linfocito *B* estimulado por un antígeno.

En nuestro primer ensayo escogimos eritrocitos de oveja como inmunógenos, ya que el anticuerpo contra tales células se detectaba fácilmente mediante una técnica desarrollada en 1963 por Niels Kaj Jerne y Albert Nordin. Mezclamos células de mieloma de ratón con células de bazo procedente de ratones inmunizados, en presencia de un agente promotor de la fusión celular; identificamos con éxito las células híbridas en un medio selectivo y observamos que segregaban inmunoglobulinas de ambos progenitores. Algunas segregaban anticuerpos contra los eritrocitos [*véase la ilustración de la página 39*]. Habíamos logrado, por tanto, aislar clones que segregaban una sola especie molecular de anticuerpo, clones que, además, podían mantenerse en cultivo. Por primera vez habíamos desarrollado cultivos continuos de células híbridas que segregaban un anticuerpo monoclonal de especificidad predefinida.

Tras nuestro éxito inicial en 1975, tropezamos con bastantes problemas y durante casi seis meses los experimentos nos condujeron a resultados poco satisfactorios. Köhler, quien a la sazón se hallaba en el Instituto de Inmunología de Basilea, daba cuenta también de sus dificultades. Giovanni Galfré se incorporó en aquellas fechas a nuestro laboratorio y ensayó varias modificaciones del método, en particular el uso de polietilenglicol como agente de fusión. Fue un período decepcionante, que nos permitió al menos optimizar las condiciones de cada etapa. Por fin descubrimos dónde estaba el problema fundamental: que el lote que empleábamos de uno de los reactivos era tóxico. Una vez remediado el fallo logramos, con Jonathan C. Howard y Geoffrey W. Butcher, del Agricultural Council Institute of Animal Physiology de Cambridge, una espectacular fusión celular que producía una serie de anticuerpos monoclonales contra antígenos de histocompatibilidad de rata: los marcadores celulares superficiales que establecen la identidad individual y son responsables del rechazo de los injertos. Otros resultados llegaron rápidamente cuando obtuvimos un protocolo tipo para los experimentos y desarrollamos nuevos métodos de ensayo de la secreción de anticuerpos.

El éxito de estos experimentos se agrandó merced a un hecho inesperado: de entre las células de bazo que





PRIMERA FUSION SATISFACTORIA de células de mieloma. Se realizó entre una estirpe celular de ratón que segregaba moléculas completas de IgG y un exceso de cadenas ligeras y una estirpe de rata que solamente segregaba cadenas ligeras. Mutantes de cada estirpe, deficientes, respectivamente, en los enzimas TK y HPGRT, se cultivaron conjuntamente en un medio selectivo, HAT, en el que se requerían esos dos enzimas para la supervivencia celular; solamente las células híbridas sobrevivieron y formaron clones. Los clones segregaban al medio varias moléculas híbridas, es decir, con cadenas pesadas de ratón y ligeras de rata (*izquierda*), pero no producían moléculas mixtas, compuestas por la región variable de un animal y la constante del otro (*derecha*). Estos resultados indicaron, en primer lugar, que las regiones variables y constantes se transcribían de ADN a ARN, éste maduraba y se traducía a proteína, como se indica en *b;* y, en segundo lugar, que la información genética de las células progenitoras se expresaba "codominantemente" en los híbridos. Estudios más recientes sobre el ADN han mostrado que entre los genes *V* y *C* hay secuencias intermedias, intrones, que se excluyen durante la maduración del ARN.

fusionábamos, sólo una de cada cien, aproximadamente, era una célula plasmática con actividad secretora de anticuerpos; y, sin embargo, rondaba el diez por ciento la proporción de nuestros clones híbridos que producían anticuerpos. Explicado de otra manera: obteníamos un número de híbridos positivos e inmortales diez veces mayor que lo que cabría esperar si la inmortalidad se transfiriese al azar a las heterogéneas poblaciones celulares esplénicas; según parecía, habíamos conseguido cierta selectividad además de la inmortalidad. Aunque no se ha explicado de un modo suficiente el fundamento de esta selectividad, pruebas recientes apoyan la idea de que probablemente tenga dos componentes. Por un lado, la secreción parece extenderse: linfocitos sintetizadores de anticuerpo, si bien incapaces normalmente de segregarlo, originan híbridos que no sólo sintetizan anticuerpo, sino que también lo segregan. Probablemente el progenitor mielómico proporcione la maquinaria secretora de la que carecen algunas células productoras de anticuerpo. Por otro lado, parece ser que las condiciones en que tiene lugar la fusión reducen la probabilidad de que las células del bazo que no sean linfocitos *B* originen híbridos de larga vida.

Cuando se ha establecido un clon de células fusionadas, todo el anticuerpo que segregan deriva genéticamente, por definición, de una sola célula. Pero no se trata necesariamente de un anticuerpo monoclonal en el sentido inmunológico del término, porque cada célula del clon híbrido tiene cromosomas de la célula mielómica y de la célula de bazo, y expresa ambas dotaciones cromosómicas. Potencialmente, una célula híbrida, en lugar de producir sólo los dos componentes de un anticuerpo monoclonal verdadero (un tipo de cadena ligera y otro de cadena pesada), puede elaborar dos cadenas pesadas y dos ligeras. Llamaremos a tal célula HLGK, porque segrega las cadenas pesadas y ligeras (en inglés *h*eavy y *l*ight, respectivamente) de la célula progenitora del bazo y las correspondientes cadenas gamma y kappa de la célula mielómica. Es normal en las células híbridas, particularmente en las primeras etapas de proliferación, una rápida pérdida de los cromosomas. En este caso, la pérdida detectable no es aleatoria: las cadenas pesadas (H o G) se pierden normalmente antes, y luego una u otra de las cadenas ligeras (L o K). El híbrido HLGK, por tanto, da lugar a variantes cuyos modelos de secreción son HLK o GLK, convirtiéndose posteriormente en variantes tales como HL, HK, LK, L o K.

El clon que nos interesa es el de células HL, que expresa las cadenas ligera y pesada de la célula de bazo específicamente inmunizada (aunque hay razones para conservar también otras variantes, principalmente la HK). Por consiguiente, al clonar no sólo es necesario ensayar el anticuerpo específico, sino también analizar los tipos de cadenas de la inmunoglobulina, y seleccionar un clon fuertemente secretor de HL (o HK). El proceso de selección puede simplificarse escogiendo una estirpe mutante de mieloma que sólo exprese la cadena ligera (K), y en consecuencia produzca híbridos HLK; o bien, seleccionando una que no exprese ninguna inmunoglobulina, productora desde el principio del híbrido inmunológicamente monoclonal HL.

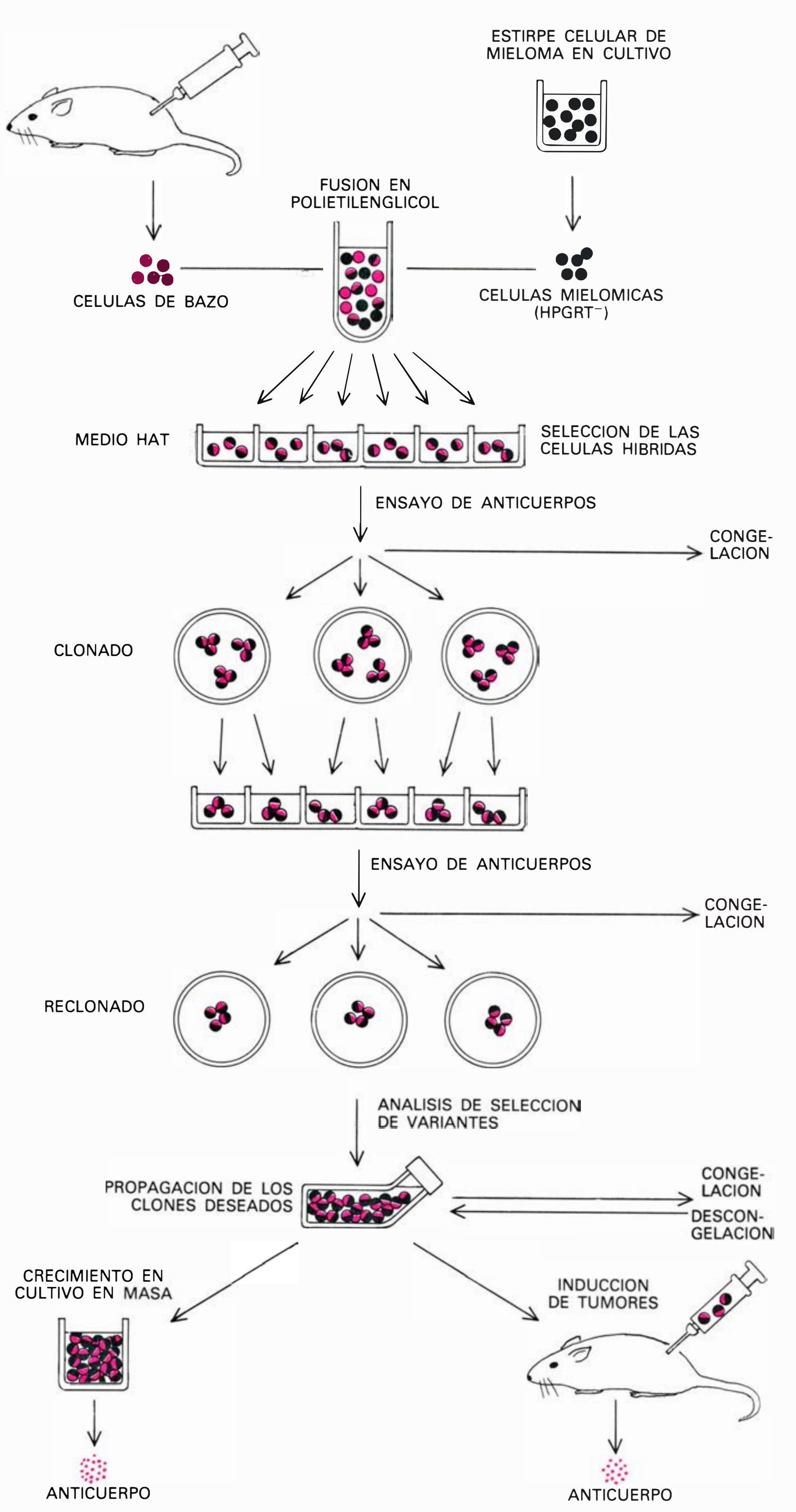
Una vez seleccionado el clon adecuado, éste puede conservarse largo tiempo congelado. En cualquier momento, puede inyectarse una muestra del clon a animales de la misma cepa que los que suministraron las células originales para la fusión. Los animales desarrollan tumores que segregan el anticuerpo monoclonal específico producido por el clon, presentándose el anticuerpo en su suero en concentraciones extremadamente altas: lo normal es más de diez miligramos de anticuerpo por mililitro de suero, alcanzando en algunos casos titulaciones de hasta un

PRIMERAS CELULAS HIBRIDAS DE MIELOMA. Se prepararon inmunizando ratones con eritrocitos de oveja (EO). Esquemáticamente se representan aquí las células implicadas en el experimento y sus inmunoglobulinas. Las células del bazo de los ratones inmunizados (que mueren en cultivo de tejidos) se fusionaron con células de mieloma de ratón deficientes en HPGRT (que mueren en el medio selectivo HAT). Las células híbridas sobrevivieron en el HAT y segregaron inmunoglobulinas procedentes de ambos progenitores. Algunas segregaban anticuerpo activo anti-EO. Se aislaron clones de células híbridas que segregaban sólo especies simples de anticuerpos frente a EO: IgG1, IgA2b e IgM (un pentámero). En otras palabras, cada clon segregaba un único anticuerpo contra los eritrocitos de oveja.

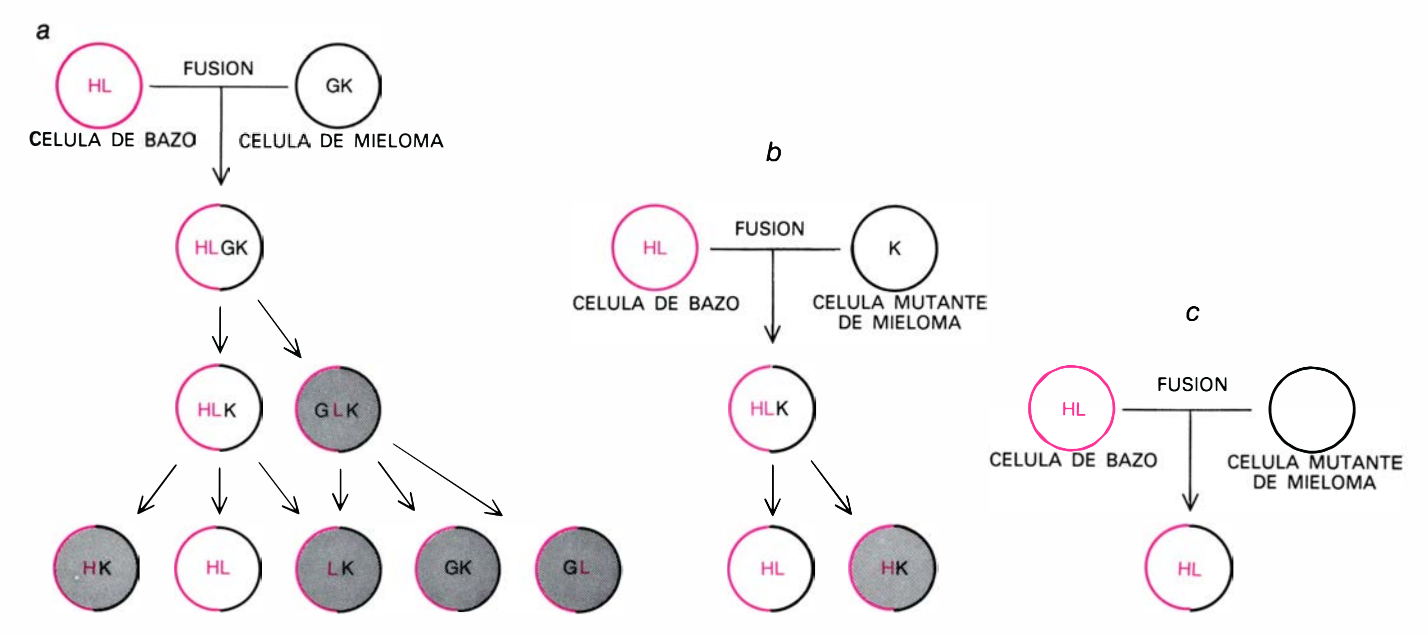
millón. Por otro lado, puede sembrarse una muestra del clon en un cultivo industrial y recolectarse el anticuerpo a partir de ese medio.

Hemos seguido métodos similares al descrito para producir anticuerpos contra un amplio repertorio de sustancias: antígenos pequeños (haptenos), proteínas (enzimas incluidos), carbohidratos y glucolípidos, componentes de las superficies celulares y virus. Los métodos parecen ser de aplicación general. Si un animal fabrica un cierto anticuerpo, podríamos inmortalizarlo en forma de mieloma híbrido. El grado de dificultad de la obtención de un mieloma híbrido parece variar con la respuesta inmune del animal. Cuando la respuesta es muy débil, se requieren procedimientos especiales para la búsqueda de un clon híbrido productor del anticuerpo específico entre los muchos clones productores de inmunoglobulinas inespecíficas. Hemos desarrollado varias formas de ensayar un gran número de células o clones para la producción de anticuerpos; y ya sabemos cómo preseleccionar bazos enriquecidos en células productoras del anticuerpo específico. Pero aquí no hay milagros: si el animal no produce un anticuerpo, éste no puede inmortalizarse.

Un anticuerpo monoclonal es un reactivo químico, bien definido, que puede reproducirse a voluntad, en contraste con un antisuero convencional, que es una mezcla variable de especies químicas y nunca se puede reproducir cuando se agota el manantial original. Conforme vayan obteniéndose anticuerpos monoclonales, reemplazarán, sin duda, a los convencionales en muchos laboratorios clínicos y de investigación. Un ejemplo lo encontramos en el ensayo tipo de los grupos sanguíneos *A*, *B*, *AB* y *O*. Los reactivos (anticuerpos de los antígenos *A* y *B* de los eritrocitos) se obtienen convencionalmente del suero humano. Se consiguen los mejores resultados hiperinmunizando al dador del suero inyectándole eritrocitos del grupo apropiado (o, mejor, del antígeno purificado), procedimiento potencialmente peligroso; donde no se realiza tal hiperinmunización, como es el caso del Reino Unido, los reactivos tienden a ser de baja calidad. Más aún, el suero donado debe ensayarse cuidadosamente para evitar la presencia de anticuerpos no deseados, cuya actividad pudiera oscurecer la reacción anti-*A* o anti-*B*. (Esta es la razón por la cual los reactivos para el ensayo no pueden



PROCEDIMIENTO CONVENCIONAL de obtención de anticuerpos monoclonales. Comienza con la fusión, mediada por polietilenglicol, de células de bazo de un ratón (o una rata) inmunizado, y células mielómicas de ratón (o rata). Los híbridos se seleccionan en HAT, se ensaya el medio para la secreción de anticuerpos y, como precaución, se congela una porción de cada cultivo positivo. Se clonan los cultivos positivos y se someten a ensayo los clones. Los que nuevamente resultan positivos se congelan, se reclonan y se analiza la presencia de variantes de inmunoglobulina. Por último, los clones elegidos pueden conservarse congelados. Al descongelarlos, las muestras pueden bien cultivarse para la producción del anticuerpo o bien inyectarse en animales para inducir mielomas que segreguen el anticuerpo deseado.



CELULAS DE MIELOMA HIBRIDO, producidas por el proceso usual de fusión, que sintetizan, no sólo las cadenas pesadas (*H*) y ligeras (*L*) de la célula de bazo progenitora, sino también las cadenas pesada gamma (*G*) y ligera kappa (*K*) de la progenitora mielómica; por cuya razón se designan HLGK. Los híbridos tienden a perder cromosomas, de suerte que, conforme crece, el "clon" pierde algunas cadenas, como se aprecia en la ilustración. Los cultivos se analizan para seleccionar las variantes deseadas (*células no sombreadas*), hasta que se aísla una variante estable HL que sólo segrega anticuerpo específico (o, para algunos propósitos, una variante HK). La fusión con células mutantes de mieloma que sólo fabrican cadenas K (*b*), o ninguna cadena de inmunoglobulina (*c*), facilita la obtención del clon HL deseado. Este se someterá ahora al procedimiento descrito en la ilustración anterior.

obtenerse inmunizando animales de laboratorio. Los anticuerpos del suero del animal reconocerían a los antígenos de los eritrocitos humanos, eliminando por completo la distinción entre los grupos *A* y *B*.)

Hemos logrado establecer que un anticuerpo monoclonal frente a la especificidad de un grupo sanguíneo no haya de ser necesariamente de origen humano. En el Medical Research Council Laboratory of Molecular Biology, en colaboración con Lennox, Steven Sachs y otros, se está fabricando, a partir de cultivos industriales, un anticuerpo que reconoce específicamente al antígeno del grupo *A*. Douglas Voak y Jack Darnborough, del Cambridge Regional Blood Transfusion Center, han contrastado el reactivo frente a los mejores disponibles comercialmente, y han comprobado que su eficacia es similar.

La posibilidad de obtener anticuerpos contra un componente simple de una mezcla "sucia" abre un nuevo camino en la purificación de productos naturales. En cuanto nos hubimos convencido de la validez de esa premisa, nos aprestamos a hacer lo que entendíamos constituía la prueba más rigurosa. La sustancia escogida fue el interferón, notoriamente difícil de purificar y obtener en cantidades significativas. Cuando Secher y Derek C. Burker, de la Universidad de Warwick, purificaron el interferón por inmunoadsorción, las mejores preparaciones conseguidas tenían una pureza del 1 por ciento. Inmunizaron ratones con una preparación de este interferón crudo, fusionaron células de bazo de los ratones con células mielómicas también de ratón y ensayaron en los clones híbridos la producción de anticuerpo anti-interferón, suplementando el ensayo biológico de la actividad anti-interferón, poco seguro y laborioso, con pruebas de secreción de inmunoglobulinas. No sin grandes dificultades seleccionaron un clon positivo e, inyectando sus células en ratones, indujeron tumores que segregaban el anticuerpo en importantes cantidades. Unieron luego el anticuerpo a partículas de carbohidrato para preparar columnas inmunoadsorbentes. El paso de una preparación totalmente cruda de interferón por la columna supuso una purificación de orden 5000 en una sola etapa. Se halla ya en fase experimental la purificación a escala industrial.

Pueden prepararse anticuerpos monoclonales específicos contra componentes singulares de cualquier mezcla compleja; asimismo, pueden producirse cantidades ilimitadas de los distintos anticuerpos para columnas inmunoadsorbentes. Este método permite descomponer una mezcla de sustancias totalmente desconocidas en sus partes integrantes. Se inmunizan los animales con la mezcla a analizar; se obtienen clones híbridos y los anticuerpos derivados de cada clon sirven para separar los componentes de la mezcla en forma secuencial.

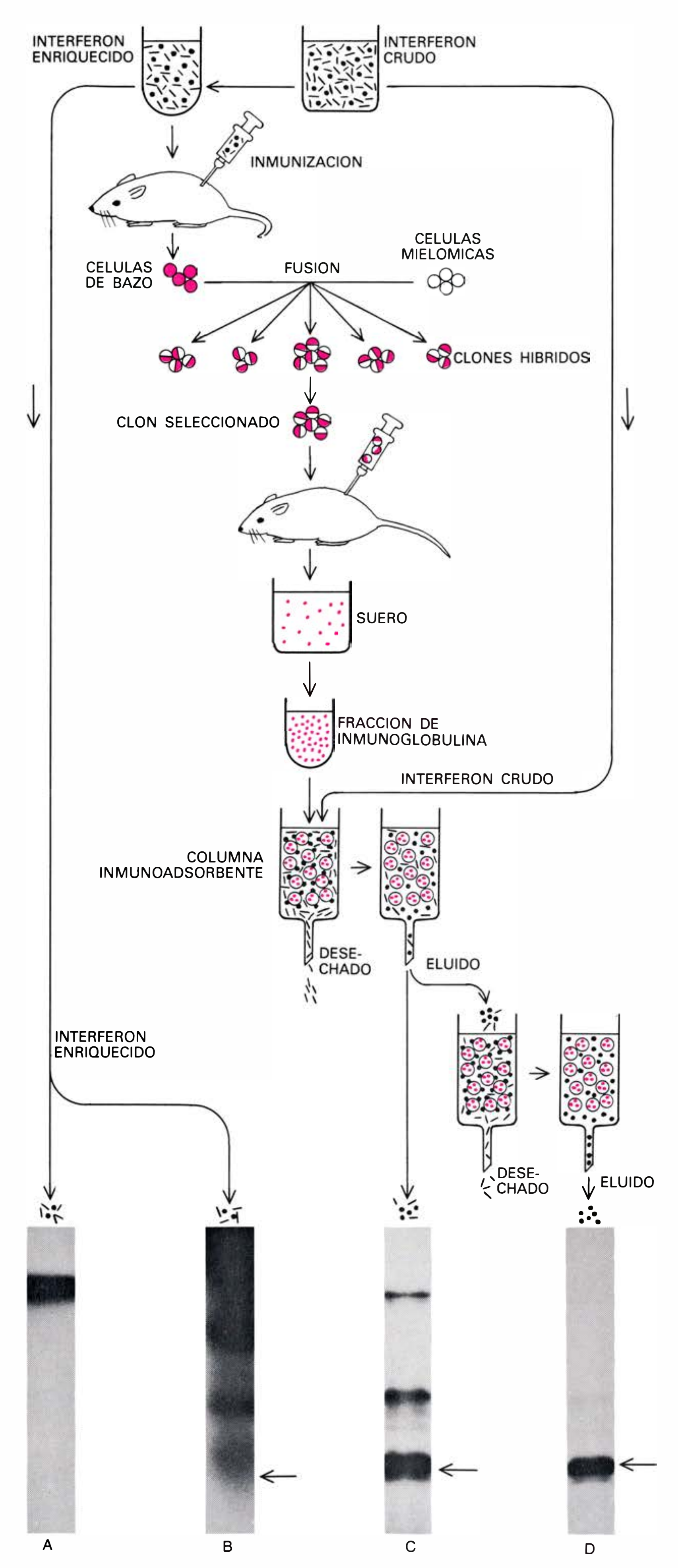
Entre las aplicaciones académicas más sugestivas de los anticuerpos monoclonales se cuenta la investigación de la biología de las membranas. Cuesta purificar las proteínas de las membranas. Se hallan en las células en pequeñas cantidades; y muchas veces su actividad biológica no es fácilmente medible, o ésta se destruye al desintegrar la membrana para su análisis. Para superar estos problemas se suele caracterizar las moléculas de las superficies celulares a través de métodos inmunológicos, enfoque este que ha resultado fructífero en el reconocimiento de los antígenos superficiales que clasifican a los distintos tipos celulares que se hallan en diferente fase de diferenciación histológica. Los anticuerpos convencionales de los antígenos superficiales son, a menudo, complejos y no reconocen moléculas individuales. Para vencer las dificultades derivadas de esta complejidad se han urgido procedimientos nada sencillos.

En 1977, Galfré y el autor, junto con Alan F. Williams, de la Universidad de Oxford, comprobaron la utilidad de la técnica del mieloma híbrido para la identificación de antígenos de diferenciación. Inmunizamos un ratón con membranas celulares de timo de rata. Una vez inmortalizadas las células de ratón que producían anticuerpos contra los linfocitos de timo pudimos aislar clones productores de distintos anti-

cuerpos específicos. En ese ensayo definimos tres nuevas características antigénicas, tarea en la que hubiéramos invertido años con los procedimientos inmunológicos convencionales. Desde entonces se han preparado muchos otros anticuerpos monoclonales contra antígenos de diferenciación de células de rata, ratón y hombre.

La estructura antigénica de una superficie celular establece la genealogía de la célula y define subconjuntos celulares. Por ejemplo, los linfocitos *B* se distinguen de los *T* (que intervienen en las reacciones de inmunidad celular y no en la secreción de anticuerpos propiamente dicha) en que los primeros tienen en su superficie moléculas de inmunoglobulina y los *T* tienen marcadores característicos tales como el antígeno Thy-1. Muchos marcadores de superficie no son, empero, específicos de una sola subclase celular. La inmunoglobulina característica de las células *B* puede estar presente en varios miembros de su linaje, de función diferente, desde las llamadas células de memoria (que responden a reinmunización con un antígeno al que el organismo ha sido expuesto previamente) hasta células plasmáticas (productoras de anticuerpos). Un estado diferenciado particular se distingue por la presencia de un conjunto específico de antígenos superficiales y la expresión cuantitativa de los mismos.

El establecimiento del perfil antigénico de cada tipo celular será lento y requerirá un buen número de anticuerpos monoclonales. Se ha comenza-



INTERFERON CRUDO, purificado por inmunoadsorción. Células de bazo de ratones, inmunizados con una preparación algo enriquecida de interferón, se fusionaron con células mielómicas. Para inducir mielomas se inyectó un clon de mieloma híbrido, seleccionado en base a su actividad anti-interferón y su secreción de inmunoglobulina; el anticuerpo purificado procedente del suero de los ratones con tumor se unió a partículas de carbohidrato, para preparar las columnas inmunoadsorbentes. Al atravesar el interferón crudo la columna, se enlazaba y quedaba retenido, mientras que los otros componentes de la mezcla cruda fluían a través de dicha columna. Se eluía luego el interferón. Un solo paso por la columna supuso un aumento de la actividad del interferón del orden de 5000 veces. La electroforesis de las proteínas (marcadas con radiactividad para facilitar su localización) visualiza los resultados (*abajo*). La preparación parcialmente enriquecida mostró una única banda (*A*), albúmina al parecer; la sobreexposición del mismo gel (*B*) permitía observar bandas menores, una de ellas (*flecha*) en la posición correspondiente al peso molecular del interferón. Después del paso del interferón crudo por la columna (*C*), había menos contaminantes y una banda intensa en la posición del interferón (*flecha*. Un segundo paso (*D*) produjo un material con una banda de interferón, única y destacada.

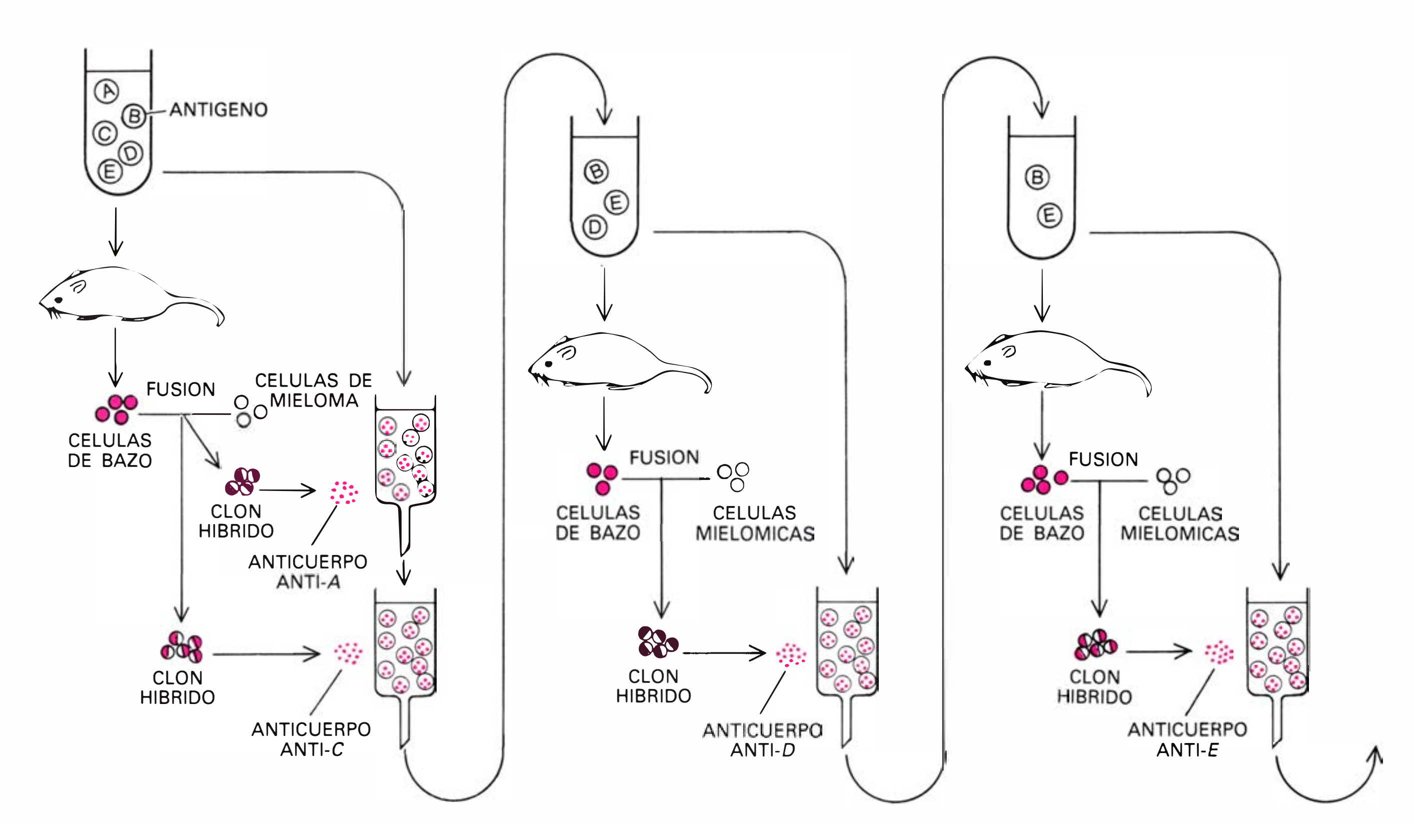
do satisfactoriamente con los pocos reactivos ya disponibles y con la ayuda de citofluorímetros y clasificadores de células activadas por fluorescencia, instrumentos que pueden medir rápidamente tamaño e intensidad de la fluorescencia de un gran número de células a las que se han unido anticuerpos monoclonales marcados con colorantes fluorescentes. De esta forma, una gran población celular puede fraccionarse en subpoblaciones en razón de su tamaño y patrón antigénico de superficie, pudiéndose estudiar a continuación la función de cada subpoblación. Los anticuerpos monoclonales son, en otras palabras, reactivos típicos que pueden identificar nuevas moléculas superficiales y, al mismo tiempo, distinguir entre diferentes poblaciones celulares. Hasta ahora, los mejores resultados se han obtenido con diversas células hemopoyéticas y linfoides; y se ha conseguido una aplicación práctica directa en el diagnóstico diferencial de distintas formas de leucemia y alteraciones afines.

El modelo de reactividad de los anticuerpos monoclonales frente a subpoblaciones de células se muestra coherente a veces con el modelo de maduración de una estirpe celular dada, pero no siempre ocurre así. Determinado antígeno monoclonal parece reconocer una característica antigénica de ciertas células de la médula ósea de la rata, en tanto que en los órganos linfoides periféricos reconoce linfocitos y, en el tejido nervioso, un componente que todavía está por identificar. Entre los linfocitos periféricos, el antígeno está presente en las células *T*, pero no en las *B*; por más que vuelva a aparecer en células plasmáticas derivadas de estas últimas. En virtud de ese fenómeno decimos que este anticuerpo monoclonal reconoce una especie de determinante antigénico "saltarín".

El enfoque monoclonal de la caracterización de antígenos de la diferenciación permite someter a examen qué antígeno se expresa en una determinada etapa y cuál es la estirpe celular que lo expresa. El método de purificación secuencial, que he descrito anteriormente, no sólo puede aplicarse a la caracterización de la complejidad antigénica de la superficie celular, sino también a la separación de componentes funcionales y estructurales de otros elementos biológicos, los orgánulos celulares y extractos celulares farmacológicamente activos, por ejemplo.

La "monoespecificidad" de los anticuerpos procedentes de clones de mielomas híbridos ha arrojado nueva luz sobre algunos fenómenos ya conocidos de las reacciones antígeno-anticuerpo. En un tubo de ensayo, la formación de un precipitado implica la reacción entre antígeno y anticuerpos convencionales. Pero ese efecto no suele observarse cuando el anticuerpo es monoclonal. Ahí reside quizá la primera prueba formal de la teoría, enunciada hace más de 40 años, de que el precipitado es una red tridimensional de antígenos y anticuerpos. Un anticuerpo monoclonal se enlaza a un solo determinante antigénico de una molécula de antígeno, por cuya razón no puede surgir una estructura reticular entre el anticuerpo monoclonal y la mayoría de los antígenos. Sólo podrá formarse la red en caso de que el antígeno sea un polímero compuesto por repetición de elementos estructurales idénticos.

La monoespecificidad ha revelado también algunos fenómenos, hasta ahora insospechados, que reclaman nuevas interpretaciones de la reacción antígeno-anticuerpo. Tomemos un ejemplo: según parece, el enlace de diferentes anticuerpos a sitios próximos del mismo antígeno es un factor importante en la ruptura de la membrana celular por el complemento. Este efecto sinérgico se descubrió por el tiempo en que nos hallábamos aislando los anticuerpos de antígenos de histocompatibilidad de la rata. Ensayábamos la



UNA MEZCLA DESCONOCIDA puede separarse utilizando los anticuerpos monoclonales provocados por ella misma. Con el conjunto aleatorio de anticuerpos producidos por animales inmunizados con la mezcla desconocida se preparan las columnas inmunoadsorbentes. Se hace pasar la mezcla por ellas, y los correspondientes antígenos quedan retenidos. La mezcla agotada, enriquecida ahora en los antígenos remanentes, se inyecta para producir más anticuerpos, y así sucesivamente. Las células híbridas proporcionan, pues, una útil herramienta para caracterizar, separar y purificar los componentes.

presencia de mielomas híbridos productores de anticuerpos midiendo la actividad citotóxica, o destructora de células de sus medios de cultivo. Los sobrenadantes de los cultivos no clonados eran consecuentemente citotóxicos; pero después de clonar las células, los sobrenadantes ya no mostraron esa capacidad. A Howard se le ocurrió medir la actividad de mezclas de sobrenadantes de estos clones aparentemente negativos. Para nuestra satisfación, la mezcla resultó activa, y nos fue fácil purificar dos componentes complementarios.

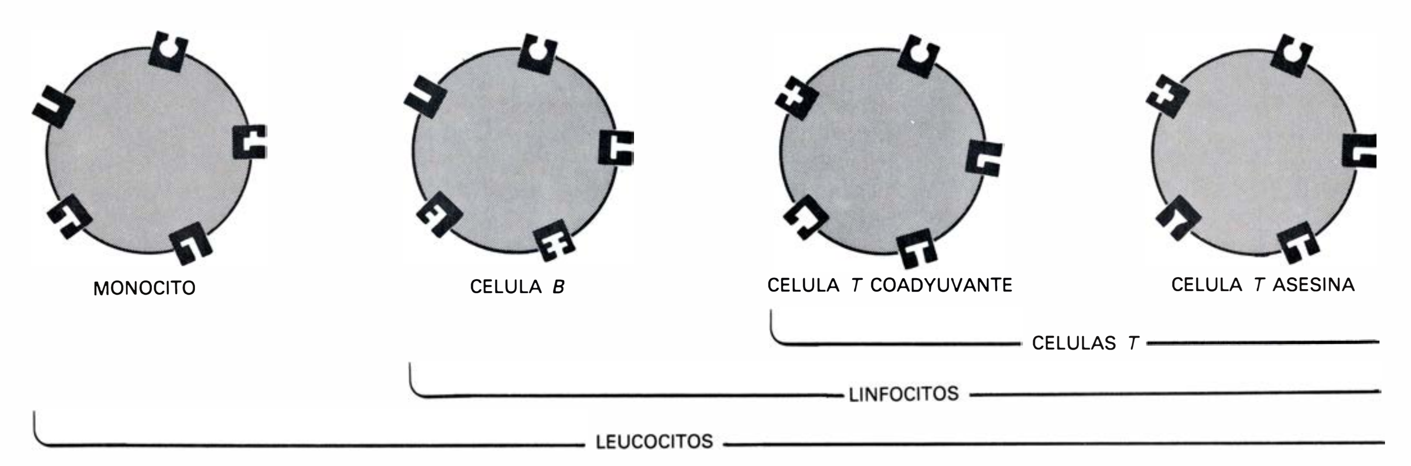
Desentrañado el efecto sinérgico, podía aprovecharse de una forma especial la actividad "muda" de los componentes aislados. Podíamos "sensibilizar" células de ensayo por exposición, primero a un anticuerpo monoclonal, y, luego, a anticuerpos de otros clones; revelando de este modo repertorios completos de anticuerpos que actuasen sinérgicamente. Hay casos evidentes en que se hace imprescindible mezclar anticuerpos monoclonales para obtener el efecto deseado. En cada ocasión habrá que decidir si las ventajas resultantes de combinar anticuerpos en proporciones específicas (en vez de confiar en las mezclas incontrolables de los antisueros ordinarios) justifica el esfuerzo necesario para producir los monoclonales.

Paulatinamente, los anticuerpos monoclonales han ido sustituyendo a los antisueros convencionales en ciertos métodos (el radioinmunoensayo, entre otros). Muchas empresas los están comercializando. Y puesto que pueden producirse en cantidades industriales, cabrá una difusión de su uso como reactivos en el diagnóstico. Tenemos un ejemplo de esto último en el anticuerpo del neurotransmisor llamado sustancia *P,* obtenido recientemente por A. Claudio Cuello y el autor. No ha hecho más que incoarse la incidencia de los monoclonales en virología, parasitología y bacteriología. Se tienen puestas grandes esperanzas en su aplicación a trasplantes de órganos, aunque esto sólo constituye un aspecto de lo que debiera ser la normalización a gran escala de la clasificación de tejidos. En investigación fundamental, las probabilidades son todavía mayores: aplicaciones, ya anunciadas, en embriología y farmacología y en el estudio de los receptores de hormonas y neurotransmisores.

Se está trabajando con todo rigor en la posible utilización de los anticuerpos monoclonales en terapia directa, cuya aplicación inmediata es la inmunización pasiva (inyección de un anticuerpo al paciente, a diferencia de la inmunización activa en que un antígeno estimula la respuesta de anticuerpos del propio paciente). Dada la impureza de los anticuerpos convencionales, la inmunización pasiva no es todavía un método generalizado de tratamiento, aunque se confía en su eficacia cuando se pueda administrar un anticuerpo purificado. En terapia tumoral se anuncian ya dos funciones para los anticuerpos monoclonales. Una tiene que ver con el mecanismo de drogas tóxicas: los anticuerpos contra los tejidos de un órgano particular o contra antígenos específicos del tumor podrían unirse a las moléculas de la droga para reforzar la acción de ésta. De un modo alternativo, podrían fabricarse anticuerpos anti-tumor que localizaran por sí mismos el tumor y atacaran a las células malignas.

Sería deseable que, para su aplicación terapéutica, los anticuerpos procedieran de linfocitos humanos, y no de ratón o rata. Contrariamente a lo que se esperaba en un comienzo, la utilización de linfocitos humanos ha resultado difícil; los intentos de inmortalizar células humanas productoras de anticuerpos, fusionándolas con células mielómicas de ratón o rata, han sido, hasta la fecha, decepcionantes. El problema reside en que, cuando se fusionan células humanas con células animales, hay una rápida pérdida preferencial de los cromosomas humanos en las células híbridas interespecíficas resultantes. Y, por ahora, no ha tenido éxito la búsqueda de una estirpe mielómica humana adecuada para el cultivo y la fusión que permita obtener un híbrido intraespecífico.

En esta panorámica general de la utilización de los anticuerpos de mielomas híbridos me he referido sólo de pasada a obvias aplicaciones en investigación inmunológica básica. Me ha parecido más interesante destacar que, si bien la técnica surgió de nuestra pretensión de desvelar la organización y expresión genética de las inmunoglobulinas, asistimos hoy a una impresionante dispersión hacia otros muchos campos. Siempre resulta difícil definir la frontera entre investigación pura y aplicada; experimentar personalmente la transición de una a otra me ha causado una profunda impresión. No puedo menos de pensar que, si el objeto de mi investigación cinco o seis años atrás hubiese sido la producción de anticuerpos monoclonales, no se me hubiera ocurrido intentar, simultáneamente, la obtención de mutantes de células secretoras de anticuerpos en un rincón del laboratorio y la fusión de dos células mielómicas en el otro; mas esa había de ser la combinación que condujo a la producción inicial de anticuerpos monoclonales contra eritrocitos de oveja.



POR ANTIGENOS DE DIFERENCIACION se entienden los determinantes antigénicos de la superficie celular que, o son específicos de tipos celulares individuales, o comunes a conjuntos o subconjuntos de tipos celulares. En la ilustración se muestran antígenos hipotéticos para cuatro leucocitos. La mejor definición de estirpe celular y células derivadas es el patrón de tales marcadores, que los anticuerpos monoclonales ayudarán a reconocer.

Ciencia y sociedad

Esperando la desintegración

Al oeste de Cleveland, Ohio, en la Morton Salt Mine, y a unos 600 metros de profundidad, se ha excavado un depósito de 18 por 21 por 24 metros. Se revestirán los lados con unos 2000 tubos fotomultiplicadores y se llenará el estanque con agua tan clara que la luz, después de atravesar la diagonal más larga del depósito, conservará todavía un tercio de su intensidad. La mayoría de los destellos de luz que se detecten se deberán a la llegada de partículas de alta energía procedentes del espacio. Pero habrá algunos que resultarán de la desintegración de un protón o de un neutrón ligado a un núcleo. Cualquiera de ellos serviría de apoyo a los esfuerzos de los físicos teóricos para establecer un formalismo matemático que unifique tres de las cuatro fuerzas básicas de la naturaleza: la fuerza fuerte, la fuerza débil y la fuerza electromagnética. (La restante, que se resiste con tenacidad, es la fuerza de la gravedad.) Constituiría también la primera observación de un proceso que lentamente va robando átomos del universo.

Al fallar en su búsqueda de tal desintegración, varios experimentos anteriores habían determinado la prodigiosa estabilidad del protón. Su vida media supera los 10^{30} años, cifra que excede la edad presente del universo en 20 órdenes de magnitud. También se cree que un neutrón ligado a un núcleo estable puede perdurar un tiempo parecido, aunque un neutrón que no esté en un núcleo tiene una vida media de minutos. Hasta hace poco se suponía que los protones y los neutrones ligados debían ser absolutamente estables. Ninguna ley de la naturaleza (o al menos ninguna ley de acuerdo con el orden fundamental) prohibía su desintegración; por otra parte tampoco se conocía ningún mecanismo capaz de convertir un protón en partículas de masa menor.

Las recientes teorías unificadoras incorporan un mecanismo de este tipo. Para empezar, se cree que el protón y el neutrón son partículas compuestas, formadas por quarks. Hay otra familia de partículas elementales, los leptones; en este grupo se incluyen el electrón, el muon y el neutrino. Los quarks y los leptones se distinguen por su diferente respuesta a las fuerzas fundamentales. Los quarks están sujetos a las cuatro fuerzas, mientras que los leptones no "sienten" la fuerza fuerte. La hipótesis de que el protón no puede desintegrarse equivale a afirmar que los quarks no pueden convertirse nunca totalmente en leptones. De modo recíproco, las teorías unificadoras establecen una conexión entre la fuerza fuerte y las fuerzas fuerte y electromagnética, postulando así un mecanismo para transmutar quarks en leptones.

El mecanismo es el siguiente. El protón consta de tres quarks. En un hipotético modo de desintegración, el protón da lugar a un positrón (un electrón cargado positivamente) y un pion neutro. A continuación se desintegra el pion, la mayoría de las veces en dos fotones o cuantos de radiación electromagnética. De esta forma desaparecen los tres quarks, creándose un leptón (el positrón) y liberándose cierta energía, que transportan los fotones.

En este proceso de desintegración actúa de intermediario una partícula hipotética designada por *X*, cuya masa equivalente se supone que alcanza una energía de 10^{14} gigaelectronvolt. Para una partícula subnuclear esta masa es enorme, comparable con la masa de un virus. En virtud de su gran masa, la partícula *X* sólo posee una efímera existencia, y no hay forma alguna de detectarla directamente. Otra consecuencia es la improbabilidad de que se cree una partícula *X* en primer lugar. En el intervalo inmediato a la gran explosión había energía disponible para crear partículas *X* en profusión, pero en el universo actual la aparición espontánea de una partícula *X* constituye un suceso de extremada rareza. Por esto es por lo que los protones se desintegran con tan poca frecuencia. A partir de las estimaciones actuales de la masa de la partícula *X*, las teorías unificadoras predicen una vida media del protón de entre 10^{31} y 10^{33} años.

La búsqueda experimental para poner de manifiesto la desintegración del protón ha de tratar de detectar la presencia de los productos más duraderos de esa desintegración, como el positrón y el pion. Para lograrlo, hay que disponer de una multitud de protones y neutrones (que colectivamente se denominan nucleones) en un lugar bien protegido de cualquier influjo exterior partículas de rayos cósmicos, por ejemplo. Los emplazamientos escogidos hasta ahora son todos subterráneos y profundos. La siguiente etapa consiste en esperar que se desintegre un nucleón.

El mayor detector de los Estados Unidos (y de todos los proyectados en todo el mundo) será el depósito instrumentado que actualmente está en construcción en la mina de sal al oeste de Cleveland. En él la multitud de nucleones serán los constituyentes de 10.000 toneladas de agua. El consorcio constructor del aparato agrupa investigadores de la Universidad de California en Irvine, de la Universidad de Michigan y del Brookhaven National Laboratory.

Un segundo detector de 1000 toneladas de agua se halla en una fase adelantada del proyecto; se ubicará en una mina de plata próxima a Salt Lake City, Utah, por un consorcio de las universidades de Harvard, Wisconsin y Purdue. En ambos experimentos, la detección de una desintegración se funda en el hecho siguiente: la velocidad de la luz en el agua es menor que en el vacío. En realidad, la velocidad de la luz en el agua es menor que la velocidad de las partículas producidas en la desintegración. Bajo estas condiciones, las partículas dan lugar a la luz conocida con el nombre de radiación de Čerenkov. (El fenómeno recuerda en muchos rasgos el del avión que al volar a una velocidad mayor que la del sonido produce un estampido sónico.)

Hay varios puntos de discrepancia entre ambos. Por un lado, el detector más pequeño, que está rodeado por una coraza de hormigón y de contadores de partículas, tendrá tubos fotomultiplicadores distribuidos por el volumen de agua a fin de recoger la luz de un modo más eficiente. Los físicos que realicen el experimento en este detector saldrán al paso del problema de los influjos exteriores recurriendo a la estrategia de hacer el aparato compacto y acorazarlo macizamente. Con todo, ambos grupos americanos suponen que las influencias exteriores pueden compensarse. En cambio dos grupos europeos tratan de colocar detectores a mayor profundidad. Un consorcio italiano está preparando un pequeño prototipo de detector para un pozo en el túnel del Mont Blanc, mientras que un grupo francés tiene planes similares para el túnel de Fréjus, 80 kilómetros al sur.

Prosiguen mientras tanto los experimentos pioneros menos pretenciosos. En uno de ellos, investigadores de la Universidad de Pennsylvania controlan 200 toneladas de nucleones en la Homestake Mine, en Dakota del Sur. En otro, un grupo del Tata Institute of Fundamental Research, de la Osaka

City University y de la Universidad de Tokyo estudiarán 150 toneladas de nucleones en el Kolar Gold Field, al sur de la India. Una tercera experiencia (ya finalizada) ha analizado sólo 25 toneladas, pero colocadas a una profundidad de unos tres kilómetros. Fueron emplazadas en una mina de oro de Africa del Sur por un grupo de investigadores de la Universidad de California en Irvine, de la Case Western Reserve University y de la Universidad de Witwatersrand.

Algunos cálculos relativos al mayor de los detectores sugieren hacia dónde pueden conducir estos esfuerzos. Unas tres veces por segundo, o 10^8 veces al año, atravesará el dispositivo experimental un muon de los rayos cósmicos, produciendo una radiación de Čerenkov intensa. La alta intensidad de la luz distinguirá este suceso de la desintegración de un protón o de un neutrón ligado a un núcleo. El problema está en que algunos muones interactúan con la roca o con la sal próximas al detector, mandando partículas al estanque de agua. Las colisiones subsiguientes con átomos mezclados en el agua pueden originar productos de la reacción que recuerdan a los de la desintegración de un nucleón. Ello justifica que no entren en los planes de los científicos cuantos sucesos ocurran en el interior del agua a menos de dos metros de distancia de las paredes.

Hay un segundo problema para el que no se tiene solución. Los neutrinos se distinguen por atravesar la materia sin interaccionar con nada; no obstante, 1000 veces al año un neutrino que entre en el volumen útil del detector colisionará con un nucleón dando lugar a un positrón y a un pion. Este es precisamente el par de partículas que con mayor probabilidad producirá la desintegración de un nucleón. El positrón y el pion se alejarán del punto de colisión formando un ángulo inferior a 120 grados. Ambos generarán radiación de Čerenkov. Para la desintegración de un protón aislado (el núcleo de un átomo de hidrógeno en el agua), el ángulo esperado entre las dos partículas es de 180 grados. Para un nucleón ligado a un núcleo el ángulo debe superar los 130 grados. Los investigadores confían en que, de las mil interacciones anuales con neutrinos, 999 puedan desecharse; en parte, midiendo el ángulo entre los dos destellos luminosos. Lo que nos deja un suceso de neutrino por año.

Suceso que es el que determina el límite de la sensibilidad del experimento. En el volumen útil del detector hay $2,5 \times 10^{33}$ nucleones. Si la vida media de un protón o de un neutrón ligado fuese de 10^{33} años, el paso de un año implicaría 2,5 desintegraciones. Se arranca de la hipótesis de que dos quintas partes de las desintegraciones lo serán en la forma que se origina un positrón y un pion, la forma de desintegración que el dispositivo detecta mejor. Por tanto, si el experimento estuviese completamente libre de error, la observación de una desintegración por año implicaría una vida media del nucleón de 10^{33} años. Un cálculo similar asocia la observación de 10 desintegraciones anuales a una vida media de 10^{32} años y de 100 desintegraciones anuales a una vida media de 10^{31} años.

A partir de los cálculos indicados resulta que, mediante este sistema, debería detectarse cualquier valor de la vida media del protón de 10^{33} años o menos. Mas, en virtud del inseparable fondo de neutrinos, el detector no alcanza esa sensibilidad. Incluso a falta de desintegraciones de protones, es de esperar un suceso al año debido a las interacciones con los neutrinos. En la práctica, pues, la observación de una desintegración aparente por año en el transcurso de dos o tres años sólo conduciría a una cota en la vida media del protón: sólo podría decirse que es mayor que 10^{33} años.

Alguien sugeriría que habría que contar con un detector mayor. Ahora bien, a medida que crece el detector, aumenta el fondo de sucesos de neutrinos. La consecuencia es clara, si se necesitan 10^4 toneladas de nucleones para establecer un mínimo de 10^{33} años, entonces se necesitarán 10^5 o aun 10^6 toneladas para buscar una vida media diez veces mayor. La investigación podría trasladarse del subsuelo profundo a las grandes profundidades marinas. Valdría equipar los depósitos de un supertanque con los instrumentos adecuados y luego dejar que el barco se hundiera.

Flujo constante

La microbiología industrial es un viejo arte que acaba de adquirir un nuevo encanto. Desde tiempos remotos se han explotado las facultades bioquímicas de diversos microorganismos, como los hongos y las levaduras, para elaborar el pan, la cerveza, el vino, el queso y muchos otros alimentos; se han aplicado también a la fabricación de medicamentos y algunos productos más. Nos encontramos ahora ante la perspectiva de modificar genéticamente los microorganismos para ampliar su abanico de posibilidades. En ese sentido, podremos inducir la síntesis bacteriana de hormonas humanas. Sin embargo, la ingeniería genética no lo es todo en el desarrollo de una tecnología bioquímica. A través de métodos más simples deben alcanzarse las condiciones óptimas para la actividad de los microorganismos. Hasta ahora, la mayoría de los procesos biológicos se han efectuado siguiendo un método de lotes relativamente simple. El rendimiento será mucho mayor si se utilizan sistemas de flujo constante.

La fabricación de cerveza puede servir de modelo del proceso en lotes. En una gran tinaja se mezclan la materia prima y los nutrientes con un cultivo de microorganismos, que crecen y se reproducen a la vez que llevan a cabo la conversión química deseada. Al cerrarse el proceso, se vacía el tonel y se da inicio a un nuevo lote. El sistema funciona excelentemente en la fabricación de cerveza, pero presenta ciertos inconvenientes cuando se afrontan procesos más complicados. Puede resultar difícil controlar las reacciones químicas de mezclas complejas. Aun así, lo que puede llegar a ser verdaderamente perjudicial es que, al terminarse un lote, los microorganismos se pierdan. Si la obtención de los microorganismos es complicada, su pérdida puede resultar costosa; es más, en determinados productos (como los antibióticos) es inaceptable la contaminación por otros organismos.

El proceso prototipo de flujo continuo es el de una refinería de petróleo. La materia prima se introduce por un extremo de la columna de reacción y el producto mana con un flujo constante por el otro. Los diversos estadios del proceso pueden separarse sin dificultad y el producto de una columna de reacción sirve de alimento para la siguiente.

La adaptación de los procesos biológicos a un método de flujo constante deberá superar un grave escollo: cómo anclar los microorganismos en un sustrato sólido. Los microorganismos deben bañarse en el líquido que fluye por la cámara de reacción pero no deben verse arrastrados por la corriente. Por otra parte, el método de anclaje no debe entorpecer la actividad bioquímica de las células. En un coloquio sobre microbiología industrial organizado el pasado año por la Royal Society de Londres, P. Dunnill, del University College de Londres, revisó los métodos de inmovilización de células en reactores bioquímicos. Sus conclusiones se han publicado ahora en la revista de la sociedad, *Philosophical Transactions.*

Hay una industria biológica que vie-

ne utilizando desde hace tiempo el método de flujo continuo, a saber, la del tratamiento de aguas residuales. En el método más común de tratamiento, las aguas residuales atraviesan con gran lentitud un lecho de piedras, en el que las bacterias y otros microorganismos descomponen la materia orgánica. Las bacterias se adhieren a las piedras por medio de una capa de limo.

Las demás aplicaciones de la microbiología industrial suelen requerir cultivos seleccionados con más esmero. Un proceso para el que se ha desarrollado una técnica con células inmovilizadas es el de la manufactura de edulcorantes a partir de almidón. El almidón es un polímero del azúcar glucosa, y puede endulzarse por medio de la conversión de parte de la glucosa en fructosa. La reacción está catalizada por enzimas secretados por diversas especies bacterianas. En un método empleado por la compañía británica Imperial Chemical Industries Ltd., la conversión la efectúan bacterias del género *Arthrobacter.* Las bacterias se comprimen en pastillas cilíndricas secas, con las que se llenan las columnas de reacción.

La compañía japonesa Tanabe Seiyaku ha diseñado un procedimiento, con células inmovilizadas, para convertir ácido fumárico en ácido L-aspártico. La conversión la efectúan células de *Escherichia coli.* Las células se suspenden en una solución de acrilamida, que se polimeriza y forma un gel de poliacrilamida. El gel se comporta como una matriz sólida que retiene las bacterias, aunque les permite el acceso a la materia prima y al caldo nutritivo. La actividad biológica de las células atrapadas disminuye, pero sólo ligeramente: la vida media es de 120 días.

También en proyectos de laboratorio a pequeña escala se han utilizado bacterias embebidas en geles de polímeros. Por ejemplo, se han empleado células de *Arthrobacter simplex* retenidas de esta forma para la deshidrogenación del esteroide cortisol, y células mutantes de la especie bacteriana *Pseudomonas putida* para la oxidación del benceno. En estos experimentos ha podido apreciarse que la velocidad de reacción no disminuía inicialmente; por el contrario, aumentaba constantemente hasta al cabo de 30 días. La explicación de este aumento es que las células se reproducen en los poros del gel. Cuando la velocidad comienza a declinar es porque la superpoblación de la matriz rígida provoca el aplastamiento de algunas células.

Muchas de las transformaciones químicas mediadas por microorganismos están catalizadas en realidad por un solo enzima. Una alternativa a los métodos de inmovilización de células es, por lo tanto, la extracción y purificación de enzimas y su fijación a un sustrato sólido. La conversión parcial de glucosa en fructosa se lleva a cabo a escala industrial por medio de enzimas fijados. Una planta industrial de fructosa basada en el método de flujo continuo dirigida por la A. E. Staley Company tiene una capacidad de producción de unas 450.000 toneladas por año.

Parece que también el arte de los cerveceros llegará a transformarse por medio de la tecnología de células inmovilizadas o de enzimas inmovilizados. Se ha descubierto que las células de levadura pueden sobrevivir en un gel a una concentración de hasta 1000 millones por mililitro. Si se instala un gel repleto de levadura en un reactor de flujo continuo y se suministran los nutrientes adecuados, el flujo de salida constará en un 10 por ciento de alcohol etílico.

Camino entre mares

Durante años, el gobierno de Israel ha acariciado el proyecto de construir un canal entre el Mar Mediterráneo y el Mar Muerto, aprovechando el desnivel de 400 metros, que le diera energía eléctrica. Se restablecería así, además, el equilibrio de evaporación del Mar Muerto, alterado por la pérdida de un 50 por ciento del abastecimiento de agua que le llegaba del Jordán. Las aguas del río Jordán se desviaron hacia el Negev en los años 50 y han sido pieza clave en la agricultura israelí. Se espera que el nivel del Mar Muerto descienda después de 1983, cuando termine la desviación del Yarmukh, afluente principal del Jordán. Hasta 1975, el rendimiento económico de aquel proyecto se había considerado insuficiente. Sin embargo, este mismo año, un grupo de trabajo, nombrado por el gobierno israelí en 1978 y dirigido por el famoso físico Yuval Ne'eman, ha concluido que, al aumentar el precio del petróleo, el balance se ha inclinado en favor del proyecto. Apoyado en esa base documental, el gobierno ha decidido dar luz verde al proyecto.

Según el plan actual, el canal trazará una curva desde Deir el Balah, en el Mediterráneo, hasta una zona al sudeste de Beersheba, y de allí, se dirigirá hacia el noreste, hasta el Mar Muerto, cerca de Masada. En total unos 112 kilómetros. El costo, incluyendo una planta hidroeléctrica, se ha estimado en unos 680 millones de dólares. Según Ne'eman, la electricidad generada amortizaría la inversión en menos de 30 años.

La construcción de un canal entre los dos mares se ideó ya en el siglo pasado. Los colonizadores británicos la propusieron para frenar el avance ruso hacia el Canal de Suez. El plan hidroeléctrico contó con el apoyo de Theodor Herzl, uno de los fundadores del movimiento sionista, que lo incluía entre los elementos de desarrollo del presunto estado de Israel. En la actualidad el proyecto se concibe como parte integrante de la planificación de las necesidades energéticas de la nación.

A pesar de la aprobación gubernamental, el inicio de la construcción del canal deberá aguardar a la resolución de ciertos problemas. Uno de ellos es que la ruta prevista atraviesa la franja de Gaza, zona en litigio entre Israel y Egipto. Quizá deba desviarse la ruta. Por otro lado, se teme en Jordania, que comparte con Israel el control del Mar Muerto, que el aporte de agua del Mediterráneo altere la composición química del Mar Muerto, lo que podría perjudicar a las plantas de potasa jordanas de la orilla oriental.

El cumplimiento de lo dispuesto en el artículo 24 de la vigente Ley de Prensa e Imprenta, se hacen constar, a continuación, los datos en él requeridos.

Accionistas de Prensa Científica, S. A.:
Editorial Labor, S. A.
Scientific American, Inc.

Consejo de Administración:
Presidente:
José M.ª Naharro Mora
Vicepresidente:
George S. Conn
Consejero-Delegado:
D. Francisco Gracia Guillén
Vocales:
D. Juan Miró Chavarría
D. Gerard Piel
D. Dennis Flanagan
Secretario:
D. Agustín Maria Rabarté

Situación financiera al 31 de diciembre de 1979, según se deduce del balance aprobado por la Junta General ordinaria de Accionistas celebrada el día 28 de mayo de 1980

ACTIVO	
Tesorería	46.278.748,–
Existencias	6.471.378,–
Inmovilizado	943.458,–
Resultados	26.392.374,–
	80.085.958,–

PASIVO	
Capital	31.000.000,–
Cuentas acreedoras	49.085.958,–
	80.085.958,–

Las causas del color

Las causas del color son muy diversas, pero todas ellas tienen el mismo origen: es la existencia de electrones en la materia, con sus variadas respuestas a las diferentes longitudes de onda de la luz, lo que hace que el mundo sea multicolor

Kurt Nassau

¿Cuál es el motivo de que el rubí sea rojo? ¿Por qué las esmeraldas son verdes? A un nivel superficial la respuesta a estas preguntas es muy simple. Cuando la luz blanca atraviesa un rubí, emerge de él con una proporción más elevada de las longitudes de onda mayores, es decir, aquellas que para el ojo humano constituyen el color rojo. Cuando la luz atraviesa una esmeralda, la distribución de longitudes de onda es diferente, correspondiendo en este caso al color verde. Aunque esta explicación del color es fundamentalmente correcta, resulta poco satisfactoria. Se echa de menos una comprensión de la forma en que la materia altera la composición de la luz que transmite o refleja. Tanto el rubí como la esmeralda deben su color a la presencia del mismo elemento como impureza. ¿Por qué razón su color difiere de forma tan radical? ¿Cuál es el origen de las diferencias en las distribuciones espectrales que constituyen los colores de los cuerpos?

Lo que sucede es que las causas últimas del color son muy diversas. La clasificación informal que yo adoptaré aquí distingue entre unas 14 causas diferentes; algunas de estas categorías engloban fenómenos distintos, aunque relacionados entre sí. Sin embargo, con una excepción, todos los mecanismos tienen un elemento común: los colores están originados por la interacción de las ondas luminosas con los electrones. El estudio de estas interacciones ha sido motivo de preocupación central de la física durante todo el siglo xx y, en consecuencia, no debe sorprendernos que, para explicar el origen del color, sea necesario recurrir a ciertas teorías físicas fundamentales. De hecho, el color es una manifestación visible (e incluso destacada) de algunos de los efectos más sutiles que determinan la estructura de la materia.

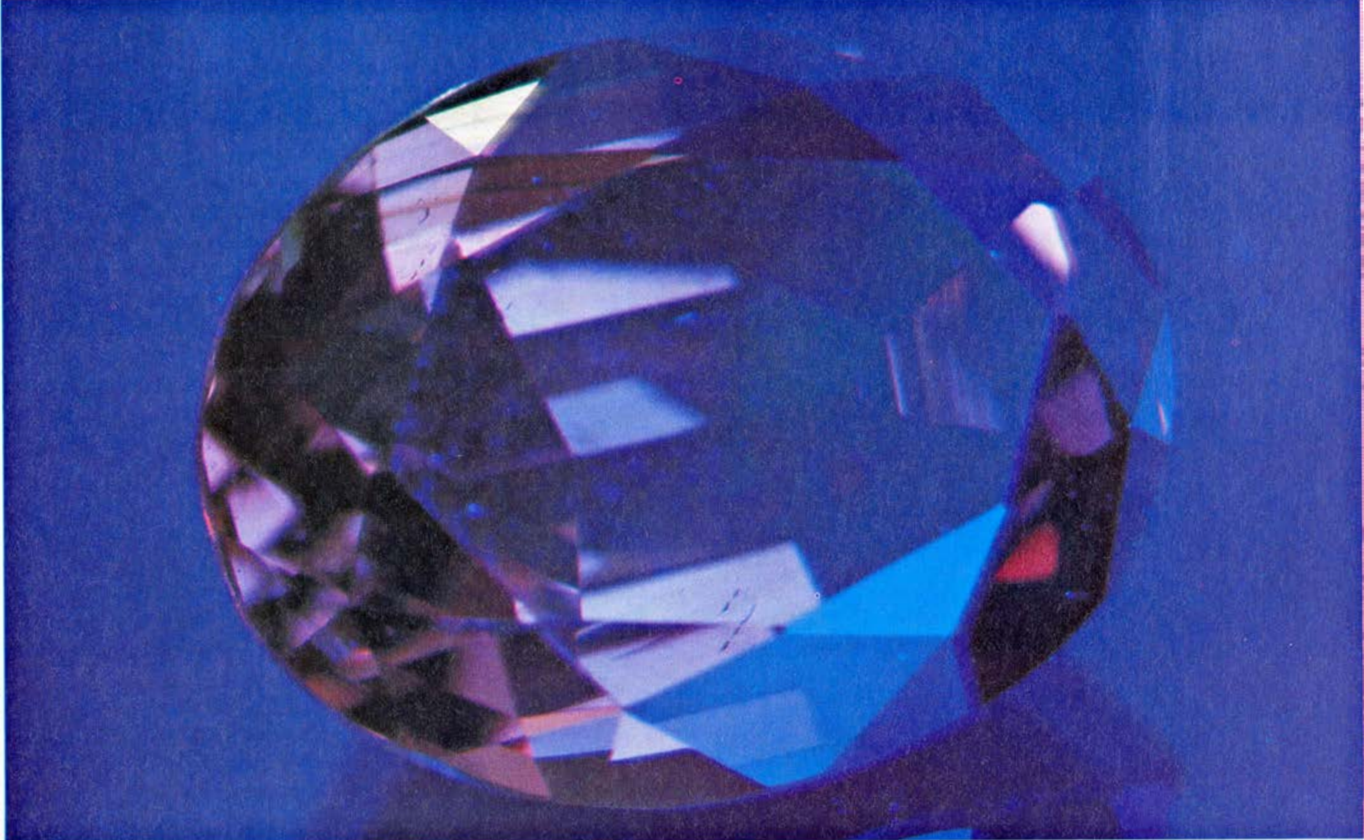
La percepción del color es una experiencia subjetiva, en la cual tanto los factores fisiológicos como psicológicos desempeñan un papel importante; estas cuestiones no serán abordadas aquí con detalle. Sin embargo, podemos suponer que el color que percibimos no es otra cosa que el resultado que proporciona la medida que lleva a cabo el ojo y la interpretación que realiza el cerebro de la longitud de onda, la frecuencia o la energía dominante de una onda luminosa. El significado de esta hipótesis está claro en el caso de una luz monocromática, cuya longitud de onda presenta un valor único y bien definido. En el caso de que la luz sea una mezcla de muchas longitudes de onda diferentes, la explicación resulta más complicada, pero son todavía las contribuciones relativas de las diferentes longitudes de onda las que determinan el color de los cuerpos.

La escalera de energía

La longitud de onda, la frecuencia y la energía son otras tantas formas alternativas de caracterizar una onda luminosa. La energía es directamente proporcional a la frecuencia; tanto la energía como la frecuencia son inversamente proporcionales a la longitud de onda. En otras palabras, frecuencias y energías altas corresponden a longitudes de onda cortas, como es el caso del violeta en el espectro visible. Una unidad utilizada frecuentemente en la medida de longitudes de onda es el nanómetro, cuyo valor equivale a 10^{-9} metros. La energía de la luz se suele medir en electronvolt; un electronvolt equivale a la energía que adquiere un electrón cuando es acelerado por una diferencia de potencial de un volt. En términos de longitudes de onda, la visión humana se extiende desde los 700 nanómetros, en el límite entre la luz roja y la radiación infrarroja, hasta los 400 nanómetros, en el límite entre la luz violeta y la radiación ultravioleta. Expresado en unidades de energía, el mismo intervalo se extiende entre 1,77 electronvolt y 3,1 electronvolt.

Todas las interacciones de la radiación electromagnética con la materia vienen determinadas por un principio de la mecánica cuántica, según el cual los átomos tan sólo pueden existir en determinados estados discretos, cada uno de los cuales se caracteriza por poseer una determinada energía; las energías intermedias a éstas están prohibidas. Cada átomo se caracteriza por una energía mínima posible, correspondiente a lo que se denomina estado fundamental, y toda una serie de estados excitados de energías progresivamente más elevadas. Los estados de energía permitida se pueden comparar con los peldaños de una escalera, aunque hay que advertir que su espaciado es altamente irregular. La luz o cualquier otra radiación únicamente puede ser absorbida en el caso de que la energía que transporta sea exactamente la necesaria para promocionar a un átomo de un peldaño a otro superior. De forma análoga, cuando un átomo sufre una transición desde un determinado estado excitado hasta otro de inferior energía, emite una radiación cuya energía será igual a la diferencia entre las energías correspondientes a los dos niveles anteriormente citados. Esta energía aparece en forma de un fotón o cuanto de luz cuya frecuencia y longitud de onda vienen determinados por dicha diferencia de energías.

Los estados de mayor interés en el análisis del color se refieren a varios niveles energéticos electrónicos. En los átomos, iones y moléculas, cada electrón ocupa un orbital determinado; cada orbital describe una distribución geométrica diferente de la carga del electrón alrededor del núcleo atómico. A su vez, los orbitales se organizan en capas. Una limitación adicional sobre los estados posibles del átomo consiste en el hecho de que cada peldaño de la escalera de energía sólo puede estar ocupado por un número limitado de electrones. En general, a medida que

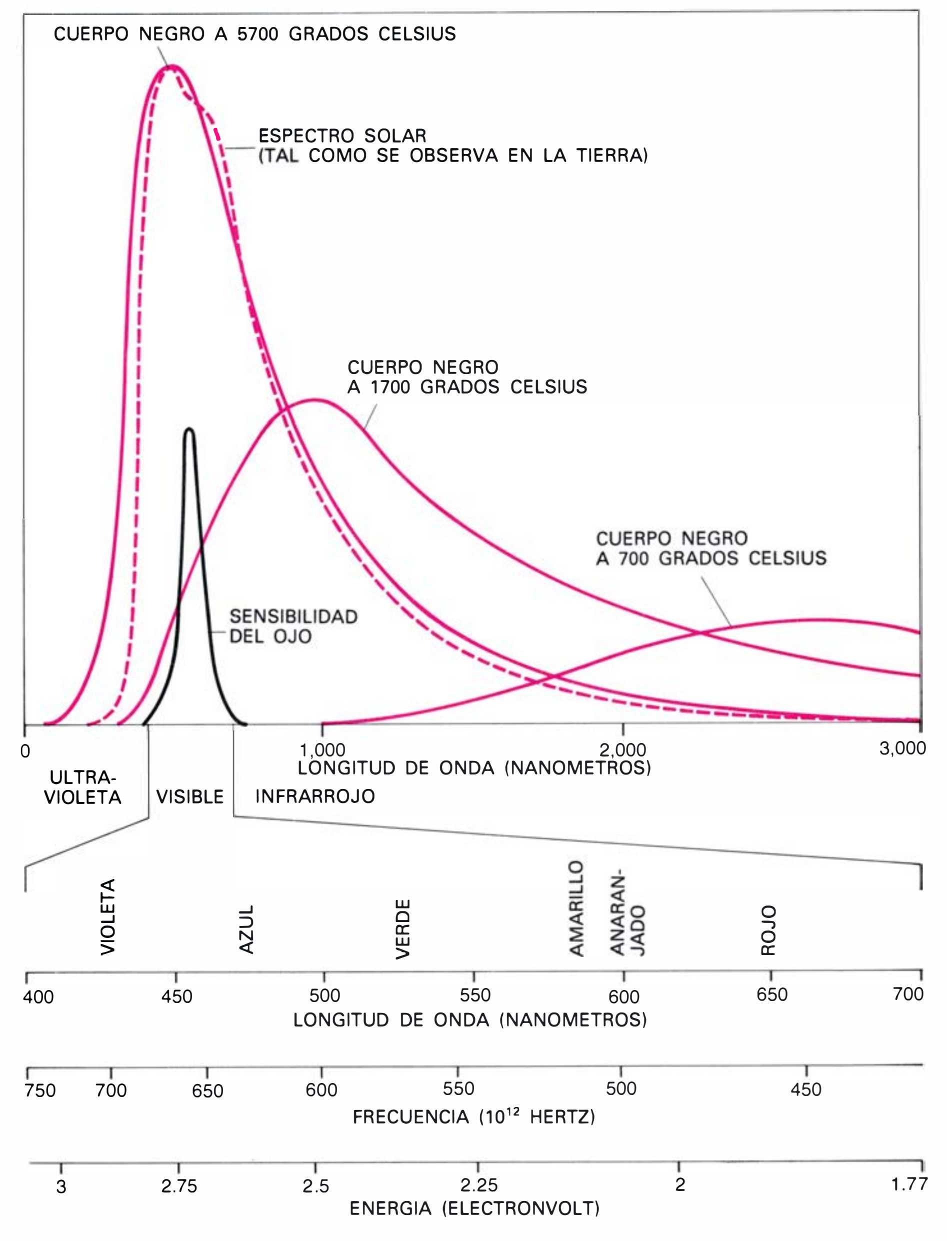


ALEJANDRITA, PIEDRA PRECIOSA que se comporta como un camaleón, adaptando su color al espectro de la luz con que se ilumina. El color de esta piedra es rojo cuando se observa a la luz de la llama de una vela o de una lámpara incandescente, rica en rojo y en amarillo, pero su color se vuelve azul verdoso bajo la luz del sol o de una lámpara fluorescente. Aunque la alejandrita es un aluminato de berilio, $BeAl_2O_4$, su color está originado por los iones de cromo presentes en el cristal como impurezas. Las transiciones entre diferentes niveles energéticos de electrones desapareados de los iones de cromo son responsables del color de muchas piedras preciosas. El color exacto depende del entorno químico en que se encuentran dichos iones. Por ejemplo, los iones de cromo dan lugar al color rojo del rubí y al color verde de la esmeralda bajo condiciones ligeramente diferentes en uno y otro caso. El entorno de los iones de cromo en la alejandrita es intermedio entre el del rubí y el de la esmeralda y, en consecuencia, su color es variable. Las alejandritas naturales son muy raras. La que se reproduce en esta figura es un cristal sintético fabricado por la empresa Creative Crystals, Inc., de Concord, California. Las fotografías no muestran los colores exactos de los cristales pues la respuesta de la película fotográfica es diferente de la del ojo humano.

avanzamos desde los átomos más pequeños hasta los mayores, los electrones se van añadiendo en forma secuencial desde el peldaño inferior hacia arriba. Dos electrones llenan la primera capa; cada una de las dos capas siguientes contiene ocho electrones. Los electrones existentes en cualquier capa completa o cerrada se aparean entre sí; la estabilidad de tales configuraciones es particularmente elevada.

Transiciones atómicas

Para promover a uno de los electrones apareados desde una capa completa hasta la siguiente vacante de la escalera se necesita una cantidad de energía relativamente grande. La energía necesaria para que tenga lugar dicha transición tan sólo puede aportarse en forma de radiación ultravioleta o incluso radiación X; como consecuencia, las capas completas no influyen de forma directa en los colores de los cuerpos. Por el contrario, el color tiene su origen en las transiciones de los electrones desapareados que, en general, son los más exteriores. Se trata precisamente de los electrones de valencia, es decir, los electrones que participan en los enlaces químicos.



ESPECTRO DE LA LUZ SOLAR. Se ajusta perfectamente a la sensibilidad del ojo. Como consecuencia, una luz nos parece blanca cuando su espectro se asemeja al solar; los otros colores se pueden describir por sus mayores o menores diferencias en relación con el espectro solar. La radiación del sol es aproximadamente igual a la de un cuerpo negro con una temperatura de 5700 grados Celsius. La forma del espectro de un cuerpo negro viene determinada totalmente por la temperatura, haciéndose más acusado y desplazándose hacia las longitudes de onda más cortas a medida que la temperatura aumenta. Así, al calentar un cuerpo, su color cambia desde el negro (ausencia de emisión) hasta el rojo, amarillo, blanco y, finalmente, azul pálido. El espectro se puede expresar en unidades de la longitud de onda, frecuencia o energía, como otras tantas formas alternativas posibles de describir una onda luminosa.

Vamos a considerar el vapor del elemento sodio, cuya densidad es lo bastante baja como para que cada átomo se comporte independientemente de sus vecinos. El átomo de sodio tiene 11 electrones, pero 10 de ellos están situados en capas completas y tan sólo el electrón de valencia toma parte directamente en las interacciones de la luz con el átomo. Cuando el átomo de sodio está en el estado fundamental, el electrón más exterior ocupa un orbital designado como $3S_{1/2}$. Los niveles energéticos inmediatamente superiores (los siguientes peldaños de la escalera) se designan por $3P_{1/2}$ y $3P_{3/2}$ y les corresponde una energía de 2,103 y 2,105 electronvolt, respectivamente, por encima del estado fundamental. Estas son las cantidades mínimas de energía que puede absorber un átomo de sodio en el estado fundamental. Las longitudes de onda correspondientes son 589,6 y 589,1 nanómetros, respectivamente, y se sitúan en la región del amarillo del espectro visible.

Por encima de los orbitales $3P$ existe una multitud de otros estados excitados, para los cuales la distancia media entre el electrón y el núcleo es mayor; lo mismo sucede con la correspondiente energía. De hecho, el número de tales estados es infinito, pero el intervalo entre los niveles se hace menor a medida que la energía aumenta, de forma que la serie converge hacia un límite finito (en el caso del sodio dicho límite vale 5,12 electronvolt). En tales condiciones, el electrón más exterior no está simplemente excitado, sino que se separa totalmente del átomo; en otras palabras, el átomo está ionizado.

Supongamos que un átomo de sodio se ioniza, quizá por efecto de un cuanto de radiación ultravioleta, y que el electrón libre y el ion se recombinan. Aun cuando el electrón puede ocupar en principio uno de los orbitales más altos, rápidamente se produce su caída hasta un nivel de energía menor. Si dicha transición tuviese lugar en una sola etapa, desde el límite de ionización hasta el estado fundamental, el átomo emitiría un fotón ultravioleta con una energía de 5,12 electronvolt. Un camino mucho más probable sería aquel que pasara a través de varios estados intermedios; en este caso, cada transición estaría acompañada por la emisión de otros tantos cuantos de energía menor. El número de cascadas posibles es limitado; ciertas "reglas de selección" determinan cuáles están permitidas. La mayoría de las transiciones permitidas que conducen al estado fundamental tienen lugar a través de alguno de los orbitales $3P$. Como consecuencia de

este hecho, los cuantos de luz amarilla con energías de 2,103 y 2,105 electronvolt se cuentan entre los emitidos. En realidad, estas dos líneas son, con mucho, las más brillantes del espectro del sodio atómico y son la causa de que el vapor de sodio excitado o ionizado emita una luz solar amarillo brillante.

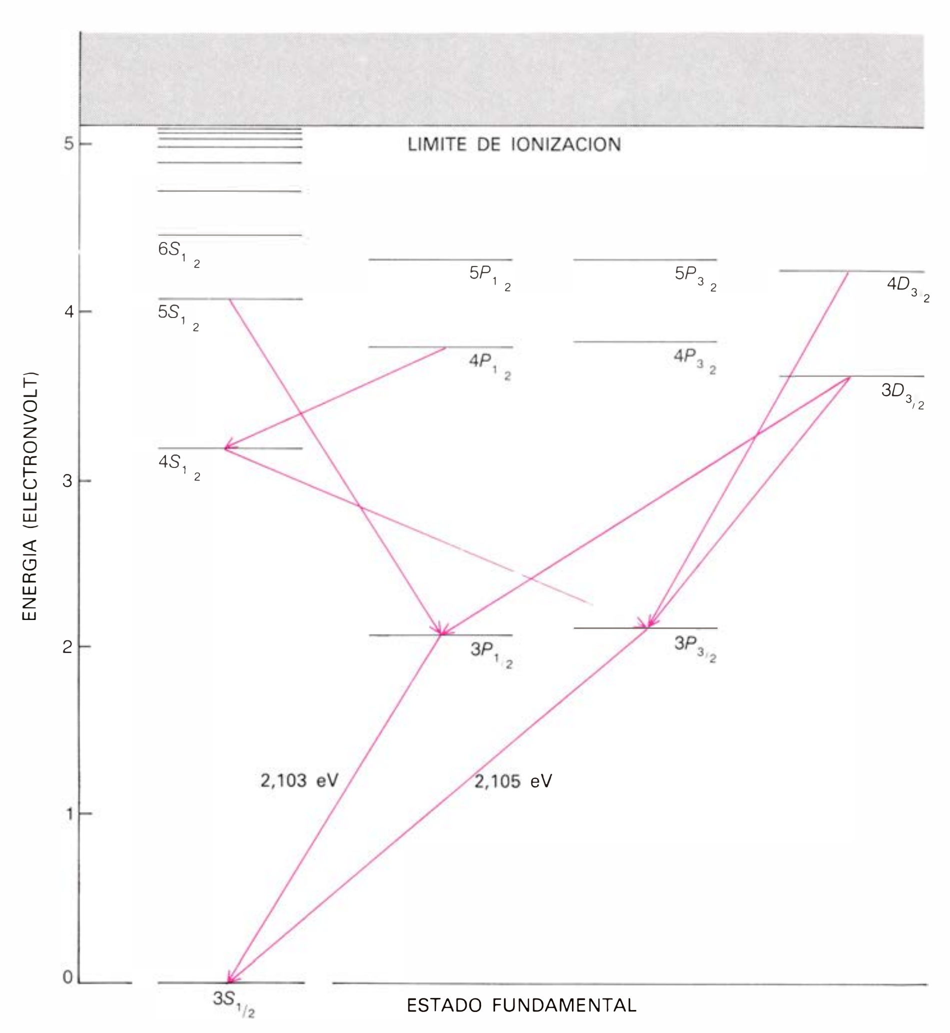
La radiación amarilla característica de los átomos de sodio se puede observar al calentar este elemento con una llama cuya temperatura sea lo suficientemente elevada como para vaporizar a algunos de sus átomos. En química analítica esta propiedad sirve de base para los ensayos a la llama a fin de detectar la presencia de sodio. El doblete de líneas amarillas es también prominente en el espectro producido por las lámparas de vapor de sodio, en las cuales los átomos de dicho elemento están ionizados como consecuencia de una descarga de alta tensión.

También otros átomos producen líneas de emisión características al ser excitados o ionizados y retornar a continuación al estado fundamental; sin embargo, el espaciado de los niveles energéticos es diferente para cada elemento y, por tanto, el color de la luz emitida en cada caso es también diferente. En el caso del neón, las líneas más intensas se sitúan en la zona espectral del rojo, lo que explica el color rojo de las luces y señales de neón. Las líneas más prominentes del átomo de mercurio se sitúan en las regiones verde y violeta del espectro visible y, por consiguiente, las lámparas de vapor de mercurio proporcionan una luz azulada que es deficitaria en rojo y amarillo. Los láseres cuyo medio de trabajo es un gas monoatómico dan lugar a líneas de emisión del tipo que corresponde al gas en cuestión. El color de los rayos y de los arcos eléctricos también tiene su origen en las excitaciones electrónicas de los átomos de determinados gases.

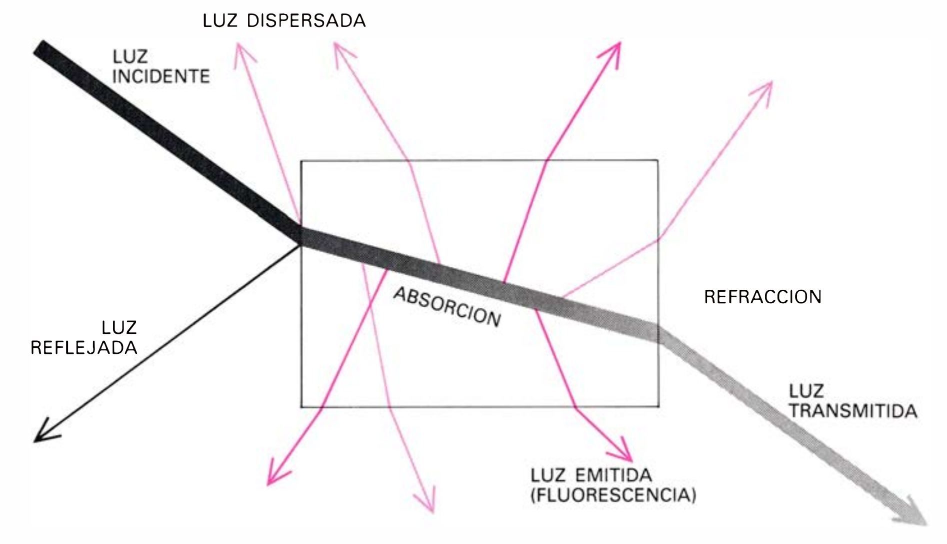
Radiación del cuerpo negro

Las líneas de emisión y absorción muy bien definidas son típicas de los gases. En general, el espectro de la luz emitida por un sólido o un líquido es muy diferente del de los gases, en el sentido de que dicho espectro se extiende a lo largo de un intervalo continuo de longitudes de onda.

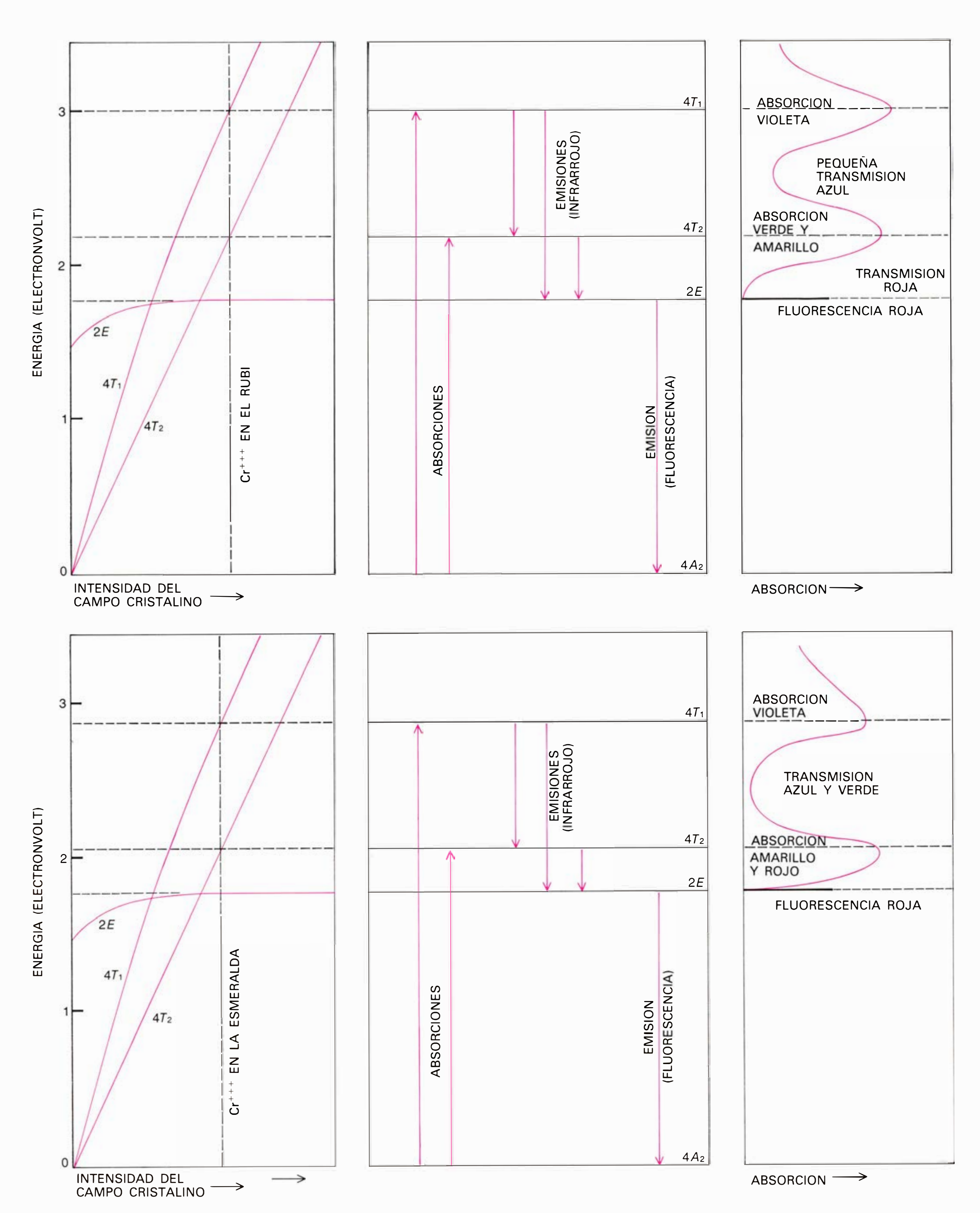
Una forma universal de la radiación que emite la materia condensada es la radiación del cuerpo negro, que constituye un espectro continuo con una forma característica. El término "cuerpo negro" se utiliza para referirnos a un material ideal que absorbe por igual a



ESTE DIAGRAMA EN FORMA DE ESCALERA corresponde al espectro de longitudes de onda del átomo de sodio. Dichas longitudes de onda son las que este átomo puede emitir o absorber. Para escalar los peldaños más altos, el átomo debe absorber un cuanto de radiación cuya energía corresponda exactamente a la diferencia de energías entre los estados inicial y final. Cuando cae a un peldaño inferior, emite un cuanto de la misma energía. La mayoría de las transiciones que conducen al estado fundamental o estado de energía más baja, designado por $3S_{1/2}$, pasan a través de los niveles $3P_{1/2}$ y $3P_{3/2}$. En estas transiciones se emiten cuantos de luz con energías de 2,103 y 2,105 electronvolt, lo cual corresponde a la región amarilla del espectro visible; por tanto, un vapor de átomos de sodio excitados presentará un color amarillo brillante. En estos diagramas en escalera sólo la dimensión vertical tiene significado, pero aquí hemos separado en sentido horizontal las series de niveles para mayor claridad de exposición.



ENTRE LAS INTERACCIONES DE LA LUZ con la materia condensada se incluyen la reflexión, la refracción, la dispersión y la absorción; una parte de la luz absorbida se puede reemitir también en forma de fluorescencia (en general, con una longitud de onda más larga). Los efectos de cada uno de estos procesos pueden variar con la longitud de onda y, en consecuencia, pueden dar lugar a la aparición de color. Por ejemplo, la absorción predominante de las longitudes de onda más cortas y la reflexión o transmisión de las más largas trae consigo que un objeto determinado sea amarillo, anaranjado o rojo. En general, la materia condensada absorbe bandas anchas y prácticamente continuas de longitudes de onda.



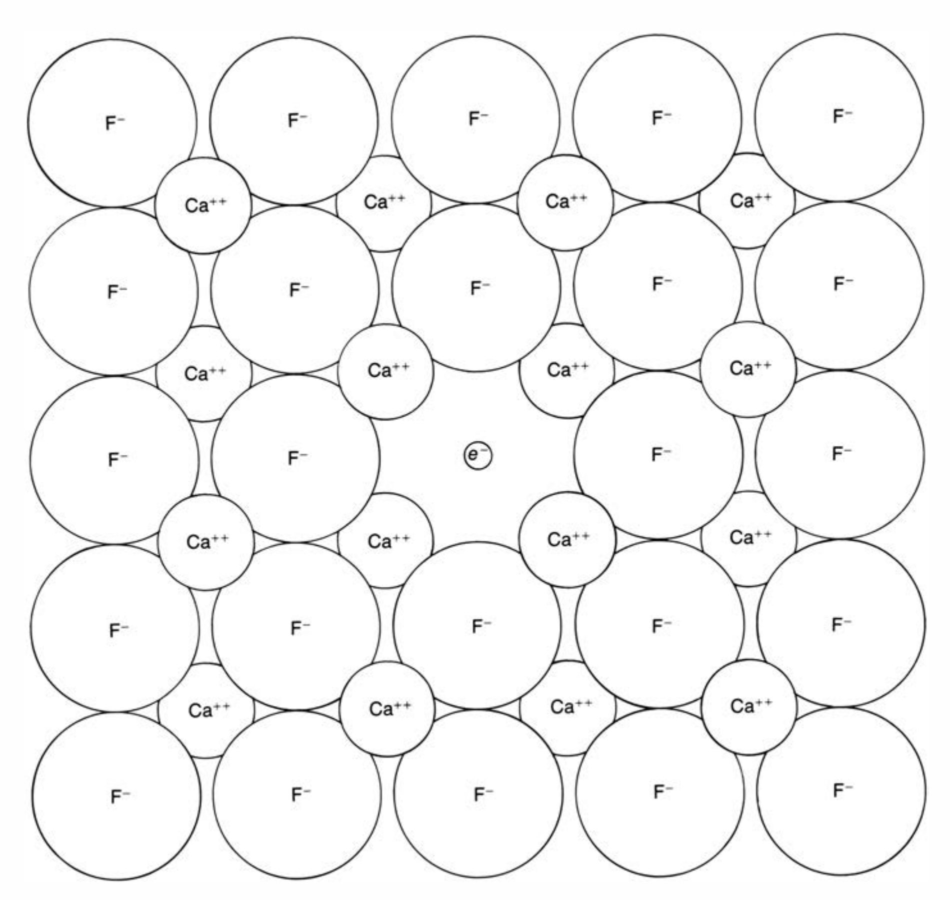
CAMPO ELECTRICO existente en el seno de un cristal. Dicho campo puede afectar su color como resultado de una alteración del estado de los átomos o iones que forman parte de la correspondiente estructura cristalina. El color del rubí o de la esmeralda es una consecuencia de la absorción de determinadas longitudes de onda por parte de electrones desapareados de los iones cromo. Las transiciones que producen dicha absorción son las mismas en ambos casos: los iones en cuestión son promovidos desde el estado fundamental, $4A_2$, a los niveles excitados $4T_2$ y $4T_1$. Las energías de estos estados vienen determinadas por la magnitud del campo cristalino. En el caso del rubí, las bandas de absorción bloquean la luz violeta, verde y amarilla; se transmite la luz roja y un poco de luz azul, lo cual da al rubí su color rojo intenso con un ligero matiz púrpura. El campo cristalino de la esmeralda es más débil y, como consecuencia de este hecho, la anchura de las bandas de absorción se reduce un tanto. Este hecho trae consigo la eliminación de la transmisión del rojo y el refuerzo de la del verde y azul. En ambos materiales, el regreso de los iones excitados al estado fundamental tiene lugar a través de un nivel intermedio designado $2E$, cuya energía resulta poco afectada por el campo cristalino. Las transiciones desde el nivel $2E$ al estado fundamental originan una fluorescencia roja que es casi idéntica a la del rubí y la esmeralda. Esta luz roja no está en el haz incidente; se origina en el cristal.

todas las longitudes de onda y que, al mismo tiempo, es también un emisor perfecto de todas las longitudes de onda. (Todos los materiales reales poseen una emisividad menor que la del cuerpo negro, pero muchos de ellos presentan un espectro parecido al del cuerpo negro para altas temperaturas.) Dicha radiación ocupa un lugar importante en la historia de la física, puesto que fue precisamente el análisis del espectro del cuerpo negro lo que permitió a Max Planck sentar las bases de la mecánica cuántica en el año 1900. Planck encontró que la forma del citado espectro tan sólo podía ser explicada sobre la base de suponer una cuantización de la energía.

En la radiación del cuerpo negro ideal, el espectro es independiente de la composición química del emisor y viene determinado por un único parámetro: la temperatura. En el cero absoluto todos los átomos ocupan el nivel disponible de menor energía y no se emite ninguna radiación. A medida que la temperatura aumenta, algunos de los átomos presentes son promovidos a determinados estados excitados, pero este proceso es de naturaleza aleatoria o estadística y los átomos se distribuyen a lo largo de un intervalo de energías muy amplio. Para cualquier temperatura finita, el número de estados ocupados aumenta gradualmente con la energía hasta alcanzar un determinado valor máximo; a partir de dicho valor, el citado número disminuye de nuevo. En consecuencia, la forma del espectro se parece al perfil de una ola del océano a punto de romper. La altura de la ola y la posición de la cresta dependen de la temperatura del cuerpo.

A temperatura ambiente, las excitaciones térmicas se hallan confinadas en la región de bajas energías y sólo se emite radiación infrarroja. Cuando la temperatura alcanza unos 700 grados Celsius, la mayoría de las emisiones corresponden todavía al infrarrojo, pero ya empieza a aparecer un poco de luz visible, que corresponde al color rojo obscuro. Si la temperatura aumenta todavía más, el máximo de la curva de emisión se desplaza hacia la región de energías más altas y longitudes de onda más cortas, de forma tal que el cuerpo se hace más brillante y su color cambia. La secuencia de los colores se desplaza desde el rojo al anaranjado, amarillo, blanco y azul pálido, lo cual está de acuerdo con términos coloquiales tales como "rojo caliente" o "rojo blanco", entre otros.

En un fuego de leña o en la llama de una vela, las partículas de carbón in-



CENTRO DE COLOR, formado cuando un electrón ocupa el lugar de un ion que ha sido desalojado de su posición en el seno de un cristal. En la fluorita o fluoruro de calcio, dicho electrón ocupa la vacante creada por la ausencia de un ion de flúor. El electrón posee un espectro de estados excitados que se sitúa en la región de las longitudes de onda visibles. Los centros de color de la fluorita, conocidos con el nombre de centros *F*, dan lugar a un color púrpura. Existen también centros de color de electrones y de "huecos", o agujeros, originados por el desplazamiento de un electrón individual en lugar de todo un ion.

candescente producen una radiación que corresponde a una temperatura de cuerpo negro del orden de 1500 grados Celsius, cuya luz se extiende del rojo al amarillo. El filamento de tungsteno de una lámpara incandescente alcanza una temperatura de 2200 grados y proporciona una luz blanco-amarillenta. La lámpara de un flash, capaz de alcanzar una temperatura de 4000 grados, proporciona una luz mucho más blanca.

El espectro solar tiene una forma muy parecida a la curva del cuerpo negro; dicha forma viene determinada por la temperatura de la superficie del sol, cuyo valor es del orden de 5700 grados C. En el espectro solar se pone de manifiesto la existencia de un máximo muy amplio centrado en 2,2 electronvolt, equivalentes a 560 nanómetros, es decir, una longitud de onda que corresponde al color amarillo-verde. La sensibilidad del ojo es máxima precisamente para esta longitud de onda. De hecho, el concepto de blanco parece estar fuertemente condicionado por el espectro de la luz diurna, que a su vez está dominado por la radiación solar. En otras palabras, la luz que percibe el ojo humano nos parece blanca cuando su espectro se asemeja al de la luz solar; de hecho, podemos definir a los otros colores en función de su diferencia respecto al espectro solar.

Colores del campo cristalino

Cuando los átomos se combinan para formar una molécula o se condensan para formar un líquido o un sólido, entran en juego nuevos modos de excitación. En particular, cabe señalar la aparición de vibraciones y rotaciones mecánicas, cuya existencia no es posible en los átomos aislados. Por ejemplo, los átomos de una molécula diatómica pueden oscilar como si estuvieran unidos entre sí por un muelle y, además, pueden girar alrededor del centro de masas de la molécula. Todos estos movimientos pueden influir ocasionalmente en el color de un material. Concretamente, la existencia de una deformación en las moléculas del agua produce la absorción de una pequeña cantidad de energía en el extremo rojo del espectro, dando lugar a la tonalidad azul característica que se observa tanto para el agua pura como para el hielo. Sin embargo, en la inmensa mayoría de los cuerpos, la energía de excitación vibracional y rotacional es muy peque-

LOS ORBITALES MOLECULARES describen la forma en que los electrones apareados se distribuyen sobre los respectivos átomos. Estos electrones suelen absorber fuertemente las longitudes de onda del espectro visible. En general, el par de electrones que constituye un enlace químico solamente absorbe en el ultravioleta, pero cuando las configuraciones alternativas de los enlaces extienden la presencia del par de electrones sobre varios átomos, la energía de excitación del par se reduce un tanto. Los orbitales moleculares que originan la producción de colores suelen hallarse asociados con los sistemas de enlaces conjugados (enlaces simples y dobles alternados) que existen en ciertos compuestos orgánicos tales como el benceno (*a*). El desplazamiento de tres pares de electrones en el benceno produce una inversión en la secuencia de los enlaces. Una forma mejor de representar la estructura de esta molécula (*b*) se basa en suponer que todos los átomos de carbono están unidos por enlaces simples y que los tres pares de electrones adicionales se distribuyen a lo largo de toda la molécula en forma de orbitales "pi". Precisamente estos orbitales pi son los responsables de la producción de color. En el caso del benceno, los estados excitados de los orbitales pi se sitúan todavía en el ultravioleta y, por consiguiente, el benceno es incoloro, pero en las moléculas con sistemas conjugados más extensos, tales como el colorante conocido con el nombre de violeta cristal (*c*), las bandas se sitúan en las longitudes de onda visibles. El color se refuerza en virtud de grupos auxocromos, que actúan como dadores y aceptadores de electrones.

ña y se disipa en forma de radiación infrarroja o calor.

Otra consecuencia de los enlaces interatómicos es el cambio que se produce en el estado de los electrones de valencia. En un átomo aislado, los electrones de valencia están desapareados y son la causa fundamental del color. Ahora bien, en las moléculas y en muchos sólidos, los electrones de valencia de los átomos forman pares con los electrones de valencia de los átomos vecinos; precisamente son estos pares los que constituyen los enlaces químicos que mantienen unidos a los átomos. Como resultado de la formación de dichos pares, las bandas de absorción de los electrones de valencia se desplazan hasta las longitudes de onda ultravioletas y ya no son aptas para la producción de color. Tan sólo aquellos electrones que permanecen en estados excepcionales pueden dar lugar a una coloración. Sin embargo, resulta evidente que tales estados excepcionales no pueden ser demasiado raros: si lo fueran, la mayoría de las moléculas y los sólidos serían transparentes a la luz visible.

En los elementos metálicos de transición, tales como el hierro, cromo y cobre, así como en los elementos del grupo de las tierras raras, aparece un conjunto de estados electrónicos excepcionales. Ciertas capas internas de los átomos de los metales en las series de transición permanecen sólo parcialmente llenas. Estas capas incompletas albergan electrones desapareados, cuyos estados excitados se sitúan frecuentemente en el espectro visible. Dichos electrones son los responsables de una amplia variedad de colores intensos. Por ejemplo, tanto el rubí como la esmeralda deben su color a que contienen pequeñísimas cantidades de cromo.

El material base del rubí es el corindón, un óxido de aluminio cuya fórmula es Al_2O_3. El corindón puro es incoloro, pero en el caso del rubí se pone de manifiesto un color muy brillante como resultado de la sustitución de un pequeño porcentaje de átomos de aluminio por iones de cromo (Cr^{+++}). Cada ion de cromo posee tres electrones desapareados, cuyo estado más bajo de energía es un estado fundamental designado por $4A_2$; existe también un espectro complejo de estados excitados. Todos los estados excitados son anchos y forman bandas; dichos estados resultan modificados por la presencia de la matriz del cristal. Aunque la identidad de los estados viene definida por la configuración electrónica del ion cromo, la posición

absoluta de cada nivel en el espectro de energías viene determinada por el campo eléctrico en que el ion está inmerso. A su vez, la simetría e intensidad del campo vienen determinadas por la naturaleza de los iones que rodean al cromo y por la disposición en que se encuentran dichos iones.

En el rubí, cada ion de cromo está situado en el centro de un octaedro deformado constituido por seis iones oxígeno. La longitud de los enlaces interatómicos en el rubí es de 0,19 nanómetros y su carácter es iónico en un 63 por ciento; esto significa que los pares de electrones que constituyen los enlaces están situados durante más tiempo cerca del oxígeno que del aluminio o del cromo. Esta distribución de la carga electrónica da lugar a la aparición de un campo eléctrico relativamente intenso, conocido como campo cristalino o campo ligando. Cuando un ion de cromo se sitúa en este campo, las energías de tres estados excitados de sus electrones desapareados corresponden al espectro visible.

Estos tres estados se designan por $2E$, $4T_2$ y $4T_1$. Las reglas de selección prohíben las transiciones directas entre el estado fundamental y el nivel $2E$, pero pueden alcanzarse ambos niveles $4T$ a partir del estado fundamental. Las energías asociadas con estas transiciones corresponden a longitudes de onda situadas en las regiones espectrales del violeta y del verde-amarillo. Dado que los citados niveles no son líneas bien definidas, sino bandas anchas, se puede absorber un amplio intervalo de longitudes de onda diferentes. En consecuencia, cuando la luz blanca atraviesa el rubí, emerge de él privada de sus componentes violeta y verde-amarilla. Esencialmente, se transmite todo el rojo, junto con algo de azul, y ello proporciona al rubí su color rojo intenso con un ligero matiz púrpura.

De acuerdo con las reglas de selección, los electrones pueden retornar desde los niveles excitados $4T$ al estado fundamental $4A_2$ solamente a través del nivel intermedio $2E$. En las transiciones iniciales entre los niveles $4T$ y $2E$ se libera una cantidad de energía muy pequeña, correspondiente a longitudes de onda infrarrojas, pero la transición desde el estado $2E$ al estado fundamental da lugar a una emisión intensa de luz roja. Deberíamos advertir que, al contrario de lo que sucede con las bandas de transmisión, esta luz roja no está presente en el haz incidente original, sino que se genera dentro del propio cristal a través de un proceso de fluorescencia. De hecho, la fluorescencia roja del rubí se puede observar con la máxima claridad cuando el cristal se ilumina con luz verde o violeta o con radiación ultravioleta. Por otra parte, la fluorescencia puede desaparecer como consecuencia de la existencia de impurezas de hierro, que se hallan presentes con frecuencia en los rubíes naturales. La luz de un láser de rubí se produce como resultado de dicha fluorescencia roja en rubíes sintéticos, libres de hierro.

El rubí y la esmeralda

La sutileza de los colores del campo cristalino se puede poner de manifiesto comparando el rubí y la esmeralda. Al igual que en el rubí, la impureza que produce el color de la esmeralda es el ion Cr^{+++} y, también en este caso, dicho ion reemplaza al aluminio en pequeñas cantidades. La similitud entre ambos materiales va todavía más lejos: en la esmeralda los iones de cromo están rodeados también por seis iones de oxígeno, formando una configuración octaédrica, y la longitud de los correspondientes enlaces es igualmente de 0,19 nanómetros. Sin embargo, en el caso de la esmeralda, la red cristalina fundamental está constituida por un silicato de aluminio y berilio, $Be_3Al_2Si_6O_{18}$. La diferencia más significativa radica en la naturaleza de los enlaces químicos, que son un poco menos iónicos; como consecuencia de este hecho, la magnitud del campo eléctrico que rodea a un ion de cromo resulta ser un tanto menor. La energía de los dos niveles $4T$ se muestra ligeramente más baja, mientras que la posición de la banda $2E$ es prácticamente la misma en ambos casos. El efecto más importante de todas estas diferencias consiste en un desplazamiento de la banda de absorción que, en el caso del rubí, bloquea la luz verde y amarilla hacia la región amarillenta y roja del espectro. La esmeralda absorbe, por tanto, la mayor parte de la luz roja, pero en cambio la transmisión del azul y del verde resulta fuertemente reforzada.

Curiosamente, la fluorescencia en la esmeralda es casi idéntica a la del rubí. Este hecho se explica en base a que la energía del nivel $2E$ resulta muy poco alterada al reducir el campo cristalino. La semejanza entre ambas emisiones fluorescentes demuestra que el hecho de que el color del rubí sea rojo y que la fluorescencia que presenta dicha gema sea también roja es mera coincidencia.

La alejandrita constituye un caso intermedio entre el rubí y la esmeralda en lo referente a su composición espectral; se trata de una piedra preciosa muy rara y, por consiguiente, de precio elevado. Una vez más el color está originado por los iones de cromo que reemplazan al aluminio; aunque, en este caso, la estructura cristalina básica es la del aluminato de berilio, $BeAl_2O_4$. El campo cristalino que determina el espectro de energías de los iones de cromo es más fuerte que en el caso de la esmeralda, pero más débil que en el caso del rubí; como resultado, las bandas de transmisión roja y verde se presentan bastante equilibradas. La casi igualdad de las dos bandas tiene una consecuencia extraordinaria: cuando se expone a la luz solar, rica en azul, el color de la alejandrita es azul verdoso, pero si se expone a la luz de una vela o de una lámpara incandescente, rica en rojo, dicha gema muestra un color rojo.

Los colores originados por el campo cristalino se ponen de manifiesto siempre que en un sólido existan iones con electrones desapareados. Los colores del aguamarina, del jade y del cuarzo cetrino tienen su origen en un mecanismo similar al descrito para el rubí y la esmeralda, aunque en estos casos el metal de transición que está presente como impureza es el hierro en lugar del cromo. Muchos compuestos y cristales en los que los metales de transición aparecen como componentes principales y no como impurezas se presentan también fuertemente coloreados. Entre los minerales naturales incluidos en esta categoría se cuentan la azurita azul y verde, la turquesa y la malaquita, cuyo color está producido por el cobre, y los granates rojos, cuyo color se debe al hierro. La mayoría de los pigmentos utilizados en la fabricación de pinturas son compuestos de metales de transición.

Centros de color

El mecanismo físico responsable de los colores del campo cristalino no se limita a los electrones existentes en los iones de los metales de transición; de hecho, no es necesario que los electrones sean una componente intrínseca de los átomos. La existencia de un electrón adicional, no ligado a ningún átomo determinado, será suficiente siempre que dicho electrón pueda ser atrapado por algún defecto estructural, tal como la falta de un ion o una impureza. La existencia de un "hueco", es decir, la ausencia de uno de los electrones de un par, puede tener el mismo efecto. Las anomalías de este tipo se conocen con el nombre de centros de color o centros *F* (inicial de

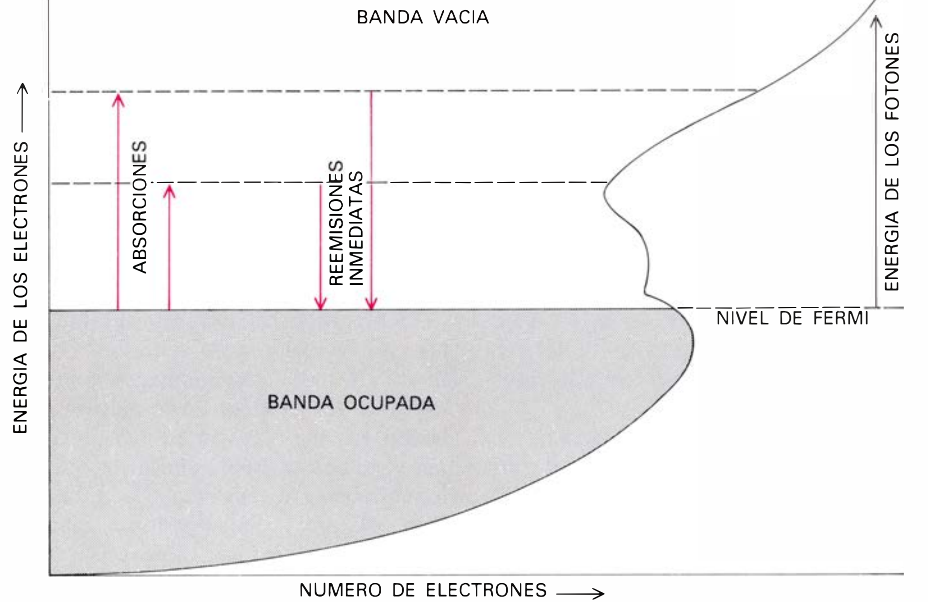
la palabra alemana *Farbe,* que significa color).

Aunque se conoce la existencia de muchas clases de centros de color, el mecanismo que produce el color tan sólo se conoce en algunos casos concretos. Uno de estos casos es el centro *F* púrpura de la fluorita, un mineral que presenta una gran variedad de centros de color. La fluorita es fluoruro cálcico, CaF_2, compuesto en el que cada ion calcio está rodeado en general por ocho átomos de flúor. Cuando desaparece un ion flúor de su posición normal, se forma un centro *F.* La pérdida de dicho ion se puede producir de diferentes formas: haciendo crecer el cristal en presencia de un exceso de calcio, exponiendo el cristal a una radiación de alta energía (que sea capaz de desplazar al ion de su posición normal) o aplicando un campo eléctrico intenso (que desplazará al flúor mediante un proceso de electrólisis). Con el fin de preservar la neutralidad eléctrica del cristal, alguna otra entidad cargada negativamente deberá ocupar el lugar que ha dejado vacante el ion de flúor en la red cristalina. En el caso de que sea un electrón el que suministre dicha carga se crea un centro *F.* El electrón queda ligado a la citada posición, no mediante un núcleo central, como sucede en los átomos o los iones, sino en virtud de la existencia del campo cristalino que crean todos los iones próximos al lugar en cuestión. En el seno de dicho campo, el electrón puede ocupar un estado fundamental y varios estados excitados, de forma análoga a lo que sucede en los metales de transición. El movimiento de los electrones entre estos estados es el origen tanto del color como de la fluorescencia.

El color del cuarzo ahumado se debe a la existencia de un centro de color constituido por un hueco. La red básica del cuarzo está constituida por dióxido de silicio (SiO_2), pero un requisito previo para la formación del centro de color es la presencia de impurezas de aluminio reemplazando a algunos iones de silicio. Dado que el aluminio tiene una valencia de + 3 y el silicio una valencia de + 4, con el fin de mantener la neutralidad eléctrica es necesaria la existencia de un ion de un metal alcalino o de un ion hidrógeno en las proximidades de dichos iones. El cuarzo suele estar contaminado con trazas de aluminio, pero la sola presencia de este elemento no daría lugar a color alguno, debido a que carece de electrones desapareados. Los centros de color aparecen cuando el cuarzo se expone durante algunos minutos a un haz intenso de rayos X o de rayos gamma o cuando ha estado sometido a niveles mucho más bajos de dichas radiaciones durante tiempos geológicos. La radiación extrae a uno de los electrones que constituyen un par de alguno de los átomos de oxígeno que están situados en las proximidades de una impureza de aluminio y, como consecuencia, en el respectivo orbital queda un electrón desapareado. El electrón ausente recibe el nombre de hueco o agujero; el electrón desapareado residual posee un conjunto de estados excitados muy parecido al que corresponde a un electrón en exceso.

La amatista es otra forma de cuarzo cuyo color se debe a un centro de color similar al descrito anteriormente, aunque en este caso la impureza es un ion de hierro en lugar de aluminio. Algunas botellas antiguas estaban fabricadas con un vidrio que contenía hierro y manganeso; después de muchos años de estar sometido a luz solar intensa, este vidrio adquiere una tonalidad púrpura como resultado de la formación de centros de color y se conoce entonces con el nombre de vidrio de amatista del desierto. El mismo efecto se consigue exponiendo dicho vidrio durante diez minutos a la acción intensa de los rayos gamma que emite el cobalto 60.

La mayoría de los centros de color son estables si el material no se calienta excesivamente. En el caso de la fluorita, una elevación de temperatura proporciona cierta movilidad al ion de flúor desplazado y, eventualmente, dicho ion puede pasar a ocupar de nuevo su posición original, con lo que el centro de color y el color desaparecen. Cuando la amatista se calienta cambia de color, convirtiéndose en cuarzo cetrino o cuarzo verdoso. Ambos colores representan la influencia del hierro sin el centro de color de la amatista. En ciertos materiales basta la propia energía de la luz solar para producir la desaparición de los centros de color. Por ejemplo, la irradiación de un topacio incoloro da lugar a la aparición de un color naranja-castaño intenso; sin embargo, el color desaparece si se expone el cristal a la luz solar durante algunos días. Por el contrario, el topacio natural del mismo color es bastante estable.



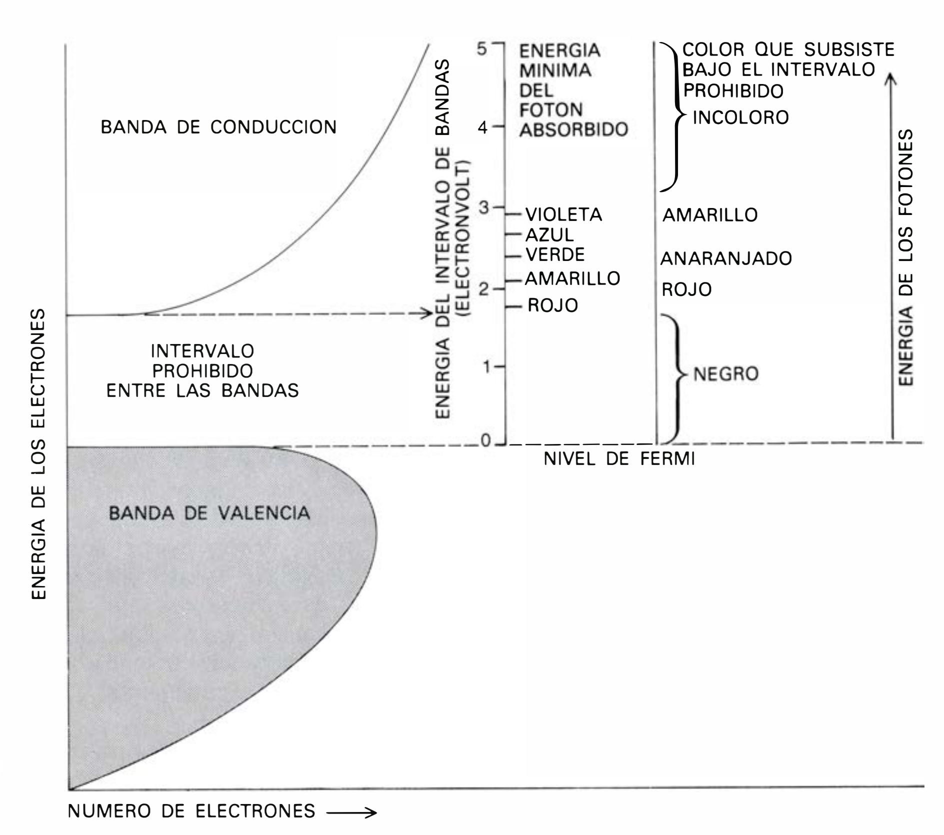
ESTRUCTURA ELECTRONICA DE UN METAL. Se caracteriza por la existencia de una banda prácticamente continua de niveles energéticos permitidos. Dicha banda está ocupada desde el estado fundamental hasta una energía denominada nivel de Fermi. Todos los niveles cuya energía es mayor que la energía de Fermi están vacíos y, por tanto, pueden aceptar electrones excitados. En virtud de esta configuración electrónica, los metales pueden absorber cualquier longitud de onda, desde el infrarrojo al ultravioleta e incluso más allá todavía. En principio, cabría esperar que el color de un material que fuera capaz de absorber luz de cualquier color sería el negro; los metales no son negros debido al hecho de que un electrón excitado puede volver inmediatamente a su estado original reemitiendo un cuanto de luz de la misma longitud de onda que el absorbido. Las superficies metálicas son, pues, altamente reflectantes.

Orbitales moleculares

Dijimos antes que los electrones de valencia de las moléculas y los sólidos están apareados en los enlaces químicos y, como consecuencia, sus estados excitados resultan desplazados hacia el ultravioleta. En realidad, esto sólo es cierto cuando los electrones apareados permanecen confinados dentro del alcance de un determinado enlace entre dos átomos. En muchos casos los electrones pueden recorrer distancias mucho mayores; de hecho, pueden mover-

se a lo largo de toda una molécula o incluso a través de un sólido macroscópico. Dichos electrones están enlazados menos fuertemente y la energía necesaria para crear un estado excitado resulta ser menor. Se dice entonces que los electrones ocupan orbitales moleculares (en contraste con los orbitales atómicos). Los orbitales moleculares son los responsables de muchos de los colores que existen en la naturaleza.

Uno de los mecanismos por los que los orbitales moleculares pueden contribuir a la aparición de color consiste en la transferencia de una carga eléctrica desde un ion a otro. Como ejemplo de este proceso podemos citar el caso del zafiro azul. Al igual que el rubí, la base del zafiro es el corindón, pero, a diferencia de aquél, el zafiro contiene dos impurezas principales, el hierro y el titanio, cuyos átomos ocupan las posiciones en que normalmente se sitúa el aluminio. En el estado de energía más baja, la carga formal del hierro es + 2 y la del titanio + 4. Cuando se transfiere un electrón desde el hierro al titanio se forma un estado excitado y la carga efectiva de ambos iones es + 3. Para que tenga lugar dicha transferencia de carga se necesita una energía de unos dos electronvolt. Tales transiciones crean una banda de absorción ancha que se extiende desde el amarillo al rojo, y proporcionan al zafiro un color azul intenso.



ESTRUCTURA DE BANDAS DE UN SEMICONDUCTOR. Es similar a la de un metal, excepto en el hecho de que una banda de energías prohibidas separa la banda de valencia ocupada de la banda de conducción vacía. En consecuencia, existe un valor mínimo en la energía que debe tener un cuanto de radiación para poder ser absorbido por estos materiales; dicha energía es precisamente igual a la energía necesaria para llevar a un electrón desde la parte superior de la banda de valencia a la parte inferior de la banda de conducción. El color de un semiconductor puro viene determinado por la anchura de este intervalo de energías prohibidas. Si corresponde al infrarrojo, todas las longitudes de onda visibles resultarán absorbidas y el material será de color negro. Si la anchura del intervalo de energías prohibidas se sitúa en la región visible, se pueden transmitir solamente unas determinadas longitudes de onda y el semiconductor se presenta coloreado con un color que va del rojo al amarillo. Cuando la energía del intervalo prohibido se sitúa en el ultravioleta, se transmiten todas las longitudes de onda visibles.

El hierro está presente en muchos materiales con sus dos valencias usuales, Fe^{++} y Fe^{+++}. La transferencia de carga entre estas dos formas da lugar a una serie de colores diferentes que, como en el caso de la magnetita, varían del azul obscuro al negro.

La teoría de los orbitales moleculares se aplica también a los colores observados en muchas sustancias orgánicas en las que los átomos de carbono (y en ocasiones los átomos de nitrógeno) se unen mediante un sistema de enlaces simples y dobles alternados, que se conocen con el nombre de enlaces conjugados. Aunque hay muchos más, el ejemplo más familiar de tales sistemas conjugados es el anillo bencénico, constituido por seis átomos de carbono. Dado que cada enlace significa un par de electrones compartidos, cuando se desplazan dos electrones desde un enlace doble al enlace simple adyacente, se invierte toda la secuencia de enlaces. Las dos estructuras definidas de este modo son totalmente equivalentes y no existe forma alguna de distinguirlas entre sí. De hecho, la mejor representación de esta estructura es la que considera que todos los átomos presentes están unidos mediante enlaces simples; los restantes pares de electrones de enlace están distribuidos a lo largo de toda la estructura en los orbitales moleculares, que en este caso reciben el nombre de orbitales pi.

El carácter extenso de los orbitales pi en un sistema de enlaces conjugados tiende a disminuir la energía de excitación de los pares de electrones. En el caso del benceno la energía del estado excitado más bajo corresponde todavía al ultravioleta y, en consecuencia, el benceno es incoloro; en el caso de moléculas mayores, y sobre todo en las que están constituidas por muchos anillos, la absorción se puede extender a la región visible. Estas moléculas orgánicas coloreadas se denominan cromóforas (palabra que significa literalmente "portadoras de color"). Con el fin de reforzar su color, se pueden unir a dicho sistema conjugado ciertos grupos químicos que sean capaces de proporcionar o aceptar electrones; estas sustancias se denominan auxocromas (literalmente, "aumentadoras de color").

Algunos pigmentos biológicos deben sus colores a sistemas extensos de orbitales pi. Como ejemplo de dichas sustancias podemos citar la clorofila verde de las plantas y la hemoglobina roja de la sangre. Los colorantes orgánicos utilizan el mismo mecanismo.

Ciertas especies químicas con orbitales moleculares extensos son también fluorescentes. Como sucede en el caso del rubí y de la esmeralda, la luz fluorescente se emite cuando la desintegración de un determinado estado excitado al estado fundamental tiene lugar a través de un nivel intermedio y la energía, de al menos una de las transiciones intermedias, corresponde a la zona visible del espectro. Los productos que refuerzan la luminosidad de los tejidos y que con este fin se suelen añadir a algunos detergentes deben sus efectos al hecho de que, a la luz del día, dichas sustancias absorben radiación ultravioleta y reemiten parte de la energía absorbida en forma de luz azul. El funcionamiento de los láseres cuyo medio activo es un colorante se basa en el mismo principio.

La fluorescencia se observa también en algunos sistemas de orbitales moleculares en los que la energía necesaria

para crear el estado excitado inicial proviene de una fuente distinta de la radiación. La bioluminiscencia de las luciérnagas y de algunos peces de las profundidades marinas está producida por una secuencia de reacciones químicas que culminan en la formación de un estado excitado de una molécula con orbitales pi extensos. Los dispositivos de "luz fría" fabricados por el hombre son una imitación de dichos procesos biológicos.

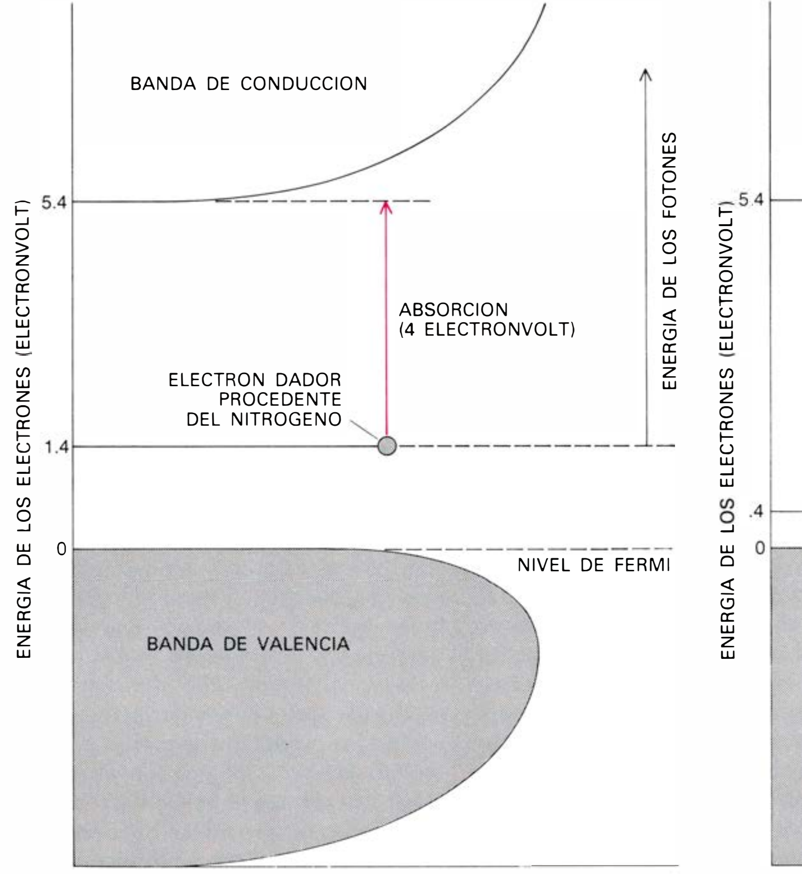
Metales y semiconductores

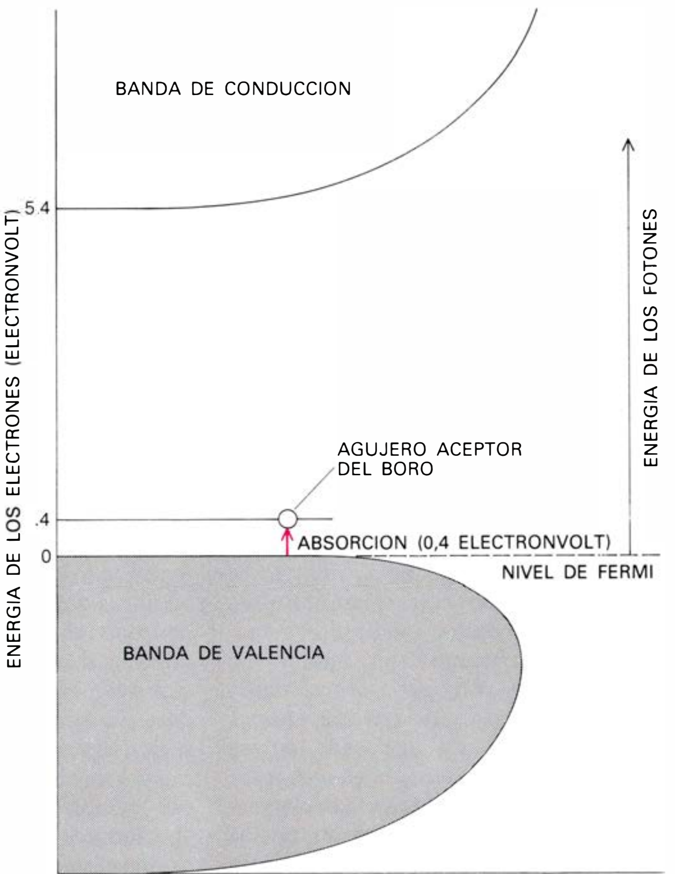
La extensión espacial de los orbitales electrónicos alcanza su máximo valor posible en los metales y semiconductores. En estos casos, los electrones se liberan totalmente de sus ligaduras a un determinado átomo o ion y se pueden mover libremente a lo largo de todo un volumen macroscópico. La distancia que pueden recorrer dichos electrones está limitada únicamente por las dimensiones del material en que se encuentran. El enorme número de electrones móviles existentes en los metales y semiconductores (del orden de 10^{23} por centímetro cúbico) proporciona a estos materiales unas propiedades ópticas y eléctricas muy singulares.

En el caso de los metales, todos los electrones de valencia son prácticamente equivalentes entre sí, puesto que pueden intercambiar libremente los lugares que ocupan. En consecuencia, se podría suponer que todos ellos poseen la misma energía, pero un principio de la mecánica cuántica prohíbe dicha posibilidad. Tanto en un sólido como en un átomo aislado, en un mismo peldaño de la escalera de energía sólo puede haber un número limitado de electrones. De acuerdo con este hecho, en un metal deben existir muchos niveles energéticos y tales niveles se presentan necesariamente muy poco espaciados entre sí. En realidad, dichos niveles energéticos forman un continuo desde el estado fundamental hacia arriba.

En el cero absoluto de temperatura, esta banda continua de estados está ocupada a partir del nivel más bajo hasta una energía conocida como nivel de Fermi; todos los estados por encima del nivel de Fermi están vacíos. Cualquier aportación de energía, no importa cuán pequeña sea, sitúa a alguno de los electrones en uno de los estados vacantes de mayor energía.

Dado que un metal posee de hecho un continuo de estados excitados, puede absorber radiación de cualquier longitud de onda. En principio, cabría esperar que una superficie que tuviera la propiedad de absorber todos los colores sería negra, pero sabemos que los metales no presentan este aspecto. Una vez más, la explicación se relaciona con la gran movilidad de los electrones metálicos: cuando uno de los electrones de un metal absorbe un fotón y salta a un estado excitado, puede reemitir inmediatamente un fotón de la misma energía y retornar a su estado original. Debido a que dicha radiación se emite muy rápidamente y con gran eficiencia, la superficie del metal se comporta como reflectante y no como absorbente, presentando el brillo característico de los metales. Si la superficie está suficientemente pulida, la re-





EXISTENCIA DE IMPUREZAS en un semiconductor. Esa presencia da lugar a la aparición de ciertos niveles de energía permitida en el intervalo de energías prohibidas. En el caso del diamante, que es un semiconductor cuyo intervalo de energías prohibidas es de 5,4 electronvolt, la presencia de impurezas de nitrógeno da lugar a la existencia de un nivel de estados ocupados situados a 1,4 electronvolt por encima del nivel de Fermi. Uno de los electrones situados en este nivel se puede excitar hasta alcanzar la banda de conducción mediante la absorción de radiación cuya energía sea de cuatro electronvolt. Dado que este nivel es relativamente ancho, en este proceso se produce cierta absorción de luz violeta y verde y el diamante muestra un color amarillo. La adición de boro a un diamante da lugar a la creación de huecos o estados vacíos en un nivel situado 0,4 electronvolt por encima del nivel de Fermi. La luz de las energías visibles más bajas puede provocar la transición de un electrón desde la banda de valencia hasta dicho nivel, con lo cual se produce la absorción de luz roja y amarilla y el diamante presenta color azul. Las impurezas alteran las propiedades eléctricas de los semiconductores.

flexión puede llegar a ser especular, como en un espejo.

Las diferencias en el color de las superficies metálicas que se ponen de manifiesto, por ejemplo, al comparar el oro con la plata, son consecuencia de las diferencias en el número de estados disponibles para una energía determinada por encima del nivel de Fermi. Dado que la densidad de estados no es uniforme, ciertas longitudes de onda son absorbidas y reemitidas con mayor eficiencia que otras. Además de la absorción y la reemisión, en algunos metales también se produce la transmisión, aunque esta última es siempre muy débil y sólo se puede observar en láminas delgadas. La transmisión varía con la longitud de onda y, por consiguiente, da lugar a la aparición de un color. Las láminas delgadas de oro, en las que la estructura normal de dicho metal está muy distorsionada, transmiten únicamente luz verde. En el caso del vidrio rubí, el oro coloidal no está sometido a tensión y la luz transmitida es de color rojo púrpura.

Colores de intervalo de bandas

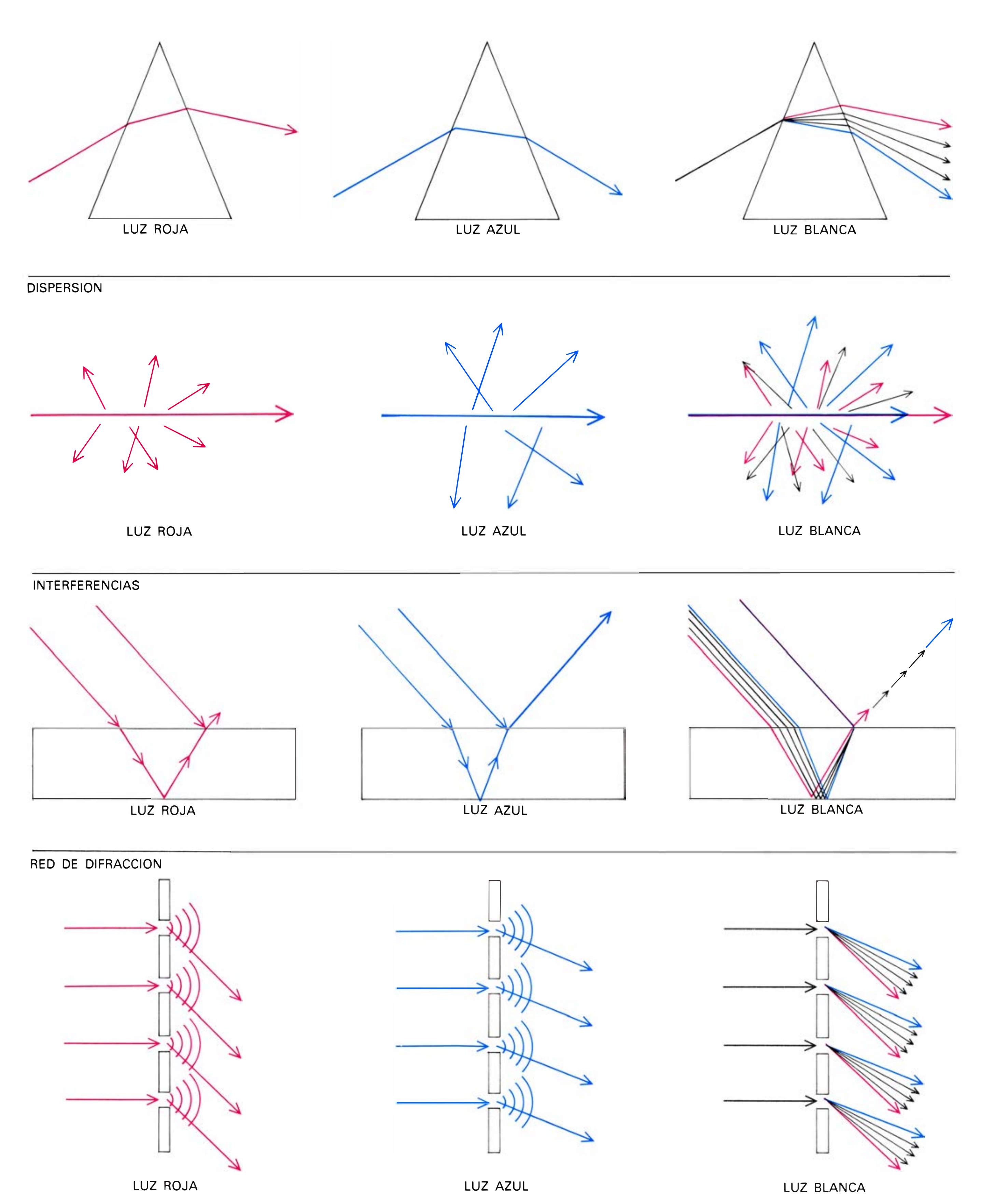
La existencia de bandas de estados electrónicos anchas y continuas es característica de otra clase de materiales: los semiconductores. En estos materiales el número medio de electrones enlazantes por cada átomo es exactamente igual a cuatro. Se incluyen aquí las formas cristalinas de algunos elementos del Grupo IV del sistema periódico, tales como el silicio, el germanio y la fase diamante del carbono; cabe señalar también la existencia de muchos semiconductores compuestos, tales como el arseniuro de galio.

Lo que distingue a un semiconductor de un metal es el desdoblamiento de la banda de niveles energéticos en dos partes. Todos los niveles de energía más baja constituyen una banda de valencia que, en el estado fundamental, está completamente ocupada. Todos los estados excitados están situados en una banda de conducción que, en el estado fundamental, está totalmente vacía. Ambas bandas están separadas entre sí por una región de energías prohibidas. La división de la banda de energías produce efectos importantes en las propiedades ópticas de los semiconductores. Los electrones ya no pueden absorber ninguna radiación cuya energía sea arbitrariamente baja; el valor mínimo de ésta es precisamente igual a la energía necesaria para elevar a un electrón desde lo más alto de la banda de valencia hasta la zona más baja de la banda de conducción.



COLORES DEBIDOS AL INTERVALO PROHIBIDO (o intervalo existente entre las bandas de valencia y conducción) de seis semiconductores diferentes. El sulfuro de cadmio puro (*abajo, a la izquierda*) tiene una energía del intervalo prohibido de 2,6 electronvolt, lo cual trae como consecuencia que sólo resulte absorbida la luz violeta; este material presenta el color complementario del violeta, es decir, el amarillo. En el seleniuro de cadmio puro (*abajo, a la derecha*), la energía del intervalo prohibido es de 1,6 electronvolt; como consecuencia, resultan absorbidas todas las longitudes de onda visibles y los cristales son negros. Los otros cuatro materiales restantes son mezclas de estos dos compuestos con proporciones diferentes de uno y otro; la correspondiente absorción presenta un gradiente, con colores que van del amarillo al rojo. Todos estos compuestos se utilizan como pigmentos bajo el nombre de rojo de cadmio.

El color de un semiconductor puro depende únicamente de la magnitud del intervalo energético que separa dichas bandas. Si el intervalo es más pequeño que la energía menor de la luz visible, todas las longitudes de onda visibles resultan absorbidas. Aquellos semiconductores en los que el intervalo de energías prohibidas es pequeño presentan un brillo metálico, como sucede con el silicio, si la reemisión de radiación es eficiente y rápida, mientras que, en caso contrario, los semiconductores son negros. Si la anchura del intervalo de energías prohibidas es mayor que la energía más alta de la luz visible, no puede producirse la absorción de las longitudes de onda visibles y



LA OPTICA FISICA proporciona la interpretación más adecuada de los colores que originan una serie de mecanismos diferentes, cuya característica común radica en la existencia de un cambio en la dirección de la luz. La dispersión de la luz blanca en sus diferentes colores componentes, producida cuando dicha luz atraviesa un prisma, es una consecuencia del hecho de que las longitudes de onda cortas se refractan según un ángulo mayor que las longitudes de onda largas. De forma análoga, la dispersión de la luz por partículas de pequeño tamaño es más efectiva para las longitudes de onda más cortas y, por tanto, la luz azul se dispersa más que la luz roja. Las interferencias se observan cuando una onda luminosa se desdobla en dos partes y éstas se reúnen de nuevo. Si las ondas en cuestión están en fase cuando se recombinan, la intensidad se refuerza; si están fuera de fase, las ondas se eliminan entre sí. Puesto que la diferencia de fase entre los dos haces depende de la longitud de onda, las interferencias pueden intensificar ciertos colores y eliminar otros. En una red de difracción la luz se difunde en muchos centros regularmente espaciados y los numerosos frentes de onda resultantes interfieren unos con otros. Cada longitud de onda es reforzada en ciertas direcciones y eliminada de otras: la luz blanca sufre una dispersión.

el correspondiente material es incoloro. El diamante, cuyo intervalo energético es de 5,4 electronvolt, constituye un ejemplo de este tipo de semiconductores; el diamante es transparente tanto a la luz visible como a un intervalo limitado de la radiación ultravioleta.

En aquellos casos en que el intervalo de energías prohibidas entre las bandas de conducción y de valencia corresponde a la luz visible, el semiconductor posee un color bien definido. El mineral de cinabrio o sulfuro de mercurio (que responde a la fórmula HgS y se conoce también con el nombre de pigmento bermellón) tiene un intervalo de energías prohibidas de 2,1 electronvolt. Todos los fotones cuyas energías son mayores que dicho valor resultan absorbidos por este mineral y solamente se transmiten las longitudes de onda visibles más largas; como consecuencia, el cinabrio presenta color rojo. El pigmento conocido como amarillo de cadmio, CdS, tiene un intervalo de energías prohibidas de 2,6 electronvolt y solamente absorbe la luz azul y violeta; el color que resulta al eliminar estas longitudes de onda de la luz blanca es el amarillo. La secuencia de los colores que origina la existencia de un intervalo prohibido entre las bandas, en orden de menor a mayor anchura del citado intervalo, es negro, rojo, anaranjado, amarillo e incoloro.

Aunque los semiconductores cuyo intervalo de energías prohibidas es grande son incoloros cuando puros, pueden aparecer coloreados cuando están "impurificados" con pequeñísimas cantidades de algún otro elemento. La impureza en cuestión puede ser un dador o un receptor de electrones; en cualquiera de ambos casos, la presencia de dicha impureza introduce un conjunto de niveles energéticos en el intervalo prohibido comprendido entre la banda de valencia y la banda de conducción. Las transiciones hacia, o desde, los nuevos niveles requieren una cantidad de energía menor que la que sería necesaria para cruzar todo el intervalo.

En el caso del diamante, al reemplazar algunos átomos de carbono por átomos de nitrógeno se crea una banda dadora situada a 1,4 electronvolt por encima de la banda de valencia. El nitrógeno posee un electrón más que el carbono, y son precisamente dichos electrones adicionales los que constituyen la banda dadora. La energía nominal necesaria para promover a uno de ellos desde la banda dadora hasta la banda de conducción es de cuatro electronvolt. Aunque esta energía corresponde todavía al ultravioleta, el nivel dador es lo suficientemente ancho como para absorber alguna luz violeta. Para una concentración de un átomo de nitrógeno por cada 100.000 átomos de carbono, el diamante es amarillo; si aumenta la proporción de nitrógeno, el color se vuelve verde.

El boro tiene un electrón menos que el carbono y, por consiguiente, la existencia de dicho elemento da lugar a la aparición de niveles aceptadores de electrones, o huecos, en el intervalo de energías prohibidas del diamante. La excitación de un electrón desde la banda de valencia puede llevarlo a ocupar uno de tales huecos o agujeros. El nivel de impurezas se sitúa en este caso a 0,4 electronvolt por encima de la banda de valencia, pero la anchura de dicho nivel es lo bastante grande como para absorber a las longitudes de onda visibles más largas. Por cuya razón, el diamante cargado con boro es de color azul.

Es un hecho bien conocido que la existencia de impurezas altera también las propiedades eléctricas de los semiconductores, haciendo posible su utilización en los numerosos dispositivos desarrollados en el marco de la electrónica del estado sólido. Entre dichos dispositivos podemos citar los diodos emisores de luz y los láseres de semiconductores. El paso de una corriente eléctrica a través de estos materiales origina la ocupación de los estados excitados; cuando los electrones retornan al estado fundamental se emite una radiación. Los semiconductores impurificados pueden funcionar también como fósforos, es decir, como materiales que emiten luz con gran eficiencia al ser estimulados eléctricamente o por otros medios. Estos fósforos son las sustancias utilizadas como fuentes luminosas en las lámparas fluorescentes y en las pantallas de los televisores. En los televisores en color se utilizan tres fósforos diferentes distribuidos regularmente sobre la superficie de sus pantallas; las longitudes de onda de la radiación que emiten dichos fósforos corresponden al rojo, verde y azul. Ciertos compuestos de los metales de transición se comportan también como fósforos.

Optica geométrica

En los fenómenos descritos anteriormente, el color es una consecuencia de la emisión directa de luz coloreada o de la absorción selectiva de determinadas longitudes de onda y la transmisión de otras. Un último grupo de fenómenos productores de color está basado en el hecho de que las interacciones de la luz con la materia cambian la dirección de los rayos luminosos. El cambio de dirección es la causa fundamental del color en la refracción y la difracción, donde la cuantía de la deflexión puede variar con la longitud de onda. En la dispersión de la luz por partículas pequeñas, la desviación que sufren los rayos no es una función determinística de la longitud de onda, pero la intensidad media de la luz difundida sí que depende de la longitud de onda. Las interferencia dan lugar a la producción de colores como resultado de la interacción de la luz con la luz; ahora bien, para que un haz luminoso sea capaz de interferir con otro es necesario un cambio en su dirección.

A un nivel muy elemental, todos estos procesos se pueden interpretar en términos de las excitaciones electrónicas de la materia. Por ejemplo, la refracción es el resultado de un cambio en la velocidad de la luz al pasar de un medio a otro; dicha velocidad viene determinada por la interacción del campo electromagnético de la radiación con las cargas eléctricas de los respectivos electrones. Siempre cabe realizar un análisis de este tipo, pero en la práctica resulta a menudo muy engorroso y proporciona muy poca información. Es preferible plantear un análisis a un "nivel superior", basándonos en los métodos de la óptica geométrica y la óptica física.

Un rayo de luz es desviado al atravesar un prisma transparente como consecuencia de que la velocidad de la luz en los sólidos es menor que en el aire. (La velocidad más alta con que se puede desplazar la luz, velocidad que según la teoría de la relatividad no puede superarse, sólo se alcanza en el vacío; en el seno de cualquier medio material, la velocidad de la luz es menor.) El hecho de que un cambio en la velocidad se traduzca en un cambio en la dirección se puede explicar si nos imaginamos un rayo luminoso como una serie de frentes de ondas planas. Cuando un frente de ondas choca oblicuamente con una superficie transparente, uno de sus extremos penetra en el correspondiente medio y es frenado antes de que el extremo opuesto alcance dicha superficie; el extremo que se frena en primer lugar queda retrasado en relación con el resto del frente de ondas y, como resultado, el rayo se desvía acercándose a la perpendicular. Al abandonar el sólido, el rayo vuelve a desviarse, pero en esta ocasión se aleja de la perpendicular.

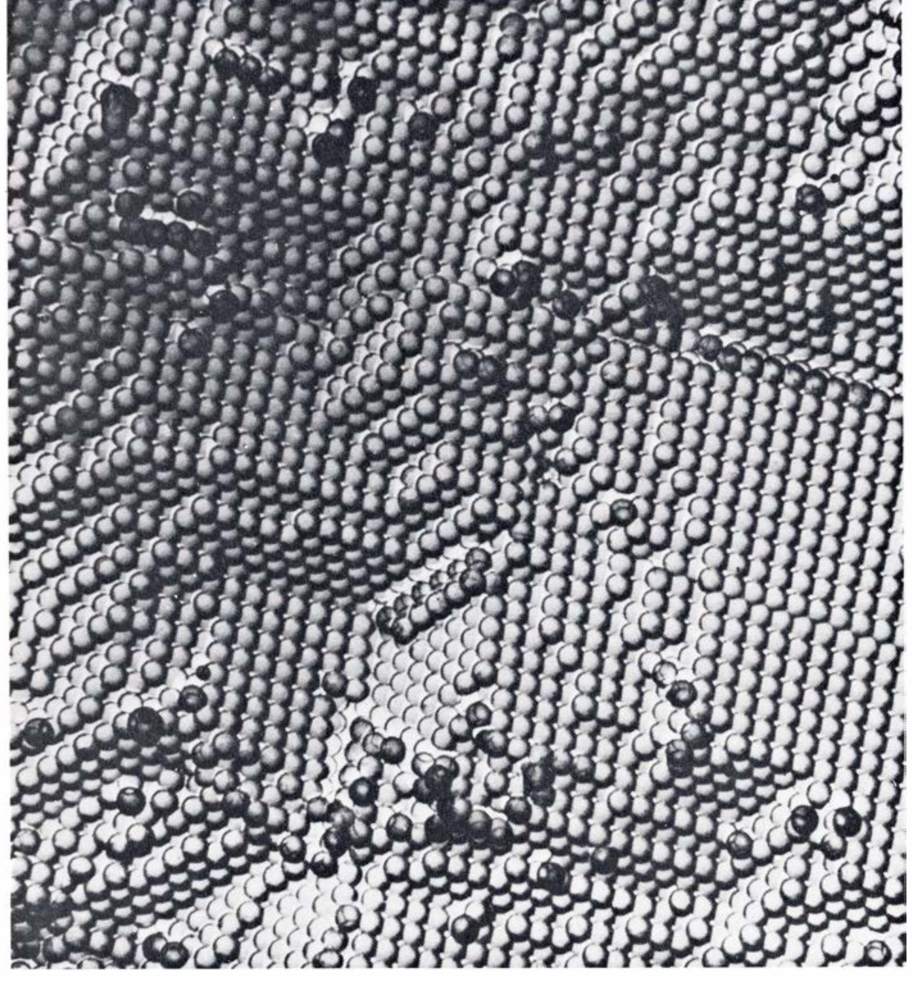
La magnitud del cambio de dirección que sufre un rayo de luz al refractarse depende del ángulo con que dicho rayo

incide sobre la superficie considerada. Para un ángulo de incidencia determinado, el cambio de dirección depende también de la relación entre las velocidades de la luz en los dos medios. En general, el valor de esta relación, conocida con el nombre de índice de refracción, no se mantiene constante al variar la longitud de onda; ésta es la razón por la que la refracción da lugar a la aparición de color. Los mecanismos que producen una disminución en la velocidad de la luz en un medio transparente resultan ser más efectivos a frecuencias altas y longitudes de onda cortas que a frecuencias bajas y longitudes de onda largas. Como consecuencia, la luz violeta resulta refractada bajo un ángulo mayor que la luz roja y cuando un haz de luz blanca atraviesa un prisma se separa en sus diferentes colores componentes.

La dispersión de la luz blanca según sus longitudes de onda fue descubierta por Isaac Newton en 1704 o, al menos, fue Newton quien reconoció su significado por primera vez. La dispersión de la luz solar por refracción en las gotas de agua o en los cristales de hielo es responsable de los colores del arco iris y de los halos coloreados que se observan en algunas ocasiones alrededor del sol y de la luna. La dispersión produce también los destellos coloreados característicos del diamante y, en menor extensión, de otras piedras preciosas transparentes. La dispersión no siempre es un fenómeno bien recibido: en los telescopios, las cámaras fotográficas y otros sistemas ópticos, da lugar a la existencia de aberraciones cromáticas, responsables de la producción de imágenes defectuosas cuando se toman en consideración objetos multicolores.

Las interferencias también fueron investigadas con detalle, por primera vez, por Newton. Este fenómeno se puede observar en un sistema de ondas en el que dos ondas se combinan entre sí para dar lugar a una nueva onda única, cuya amplitud en un punto determinado es simplemente igual a la suma de las amplitudes de las ondas originales. De este modo, una onda luminosa monocromática se puede desdoblar en dos componentes que siguen caminos diferentes y, a continuación, se combinan de nuevo. Allí donde las ondas componentes están en fase, es decir, en una situación tal en que las crestas y los valles coinciden, las ondas se refuerzan mutuamente y la luz es brillante. Allí donde las componentes están fuera de fase, las ondas se eliminan entre sí y aparece obscuridad.

Las interferencias se observan con frecuencia en las láminas transparentes delgadas, en las que una parte de la luz se refleja en la primera superficie y otra parte en la segunda. El hecho de que los haces se refuercen o se anulen depende de la naturaleza y del espesor de la lámina, así como del ángulo de reflexión y de la longitud de onda de la luz. Si el espesor de la lámina es uniforme, en cada dirección emerge una longitud de onda diferente. Si el espesor de la lámina no es uniforme, para un determinado ángulo de emergencia, en cada lugar de la lámina se observa un color diferente. Los colores que se ponen de manifiesto en ciertos escarabajos y mariposas están originados por las interferencias que se producen en alguna cutícula fina y transparente. La existencia de una capa delgada de aceite sobre la superficie del agua produce una multitud de colores como consecuencia de este mismo fenómeno.



RED DE DIFRACCION NATURAL formada por un conjunto de esferas íntimamente empaquetadas entre sí en un ópalo sintético. Las esferas están constituidas por dióxido de silicio con un poco de agua y se encuentran incluidas en el seno de una matriz transparente con una composición similar, pero con un índice de refracción diferente. El diámetro de estas esferas es del orden de 250 nanómetros. La dispersión de la luz blanca por esta red de difracción tridimensional da lugar a la aparición de una serie de colores espectralmente puros en el interior del ópalo. Este ópalo sintético ha sido fabricado por Pierre Gilson.

Las redes de difracción

La difracción consiste en la desviación que experimenta una onda luminosa al chocar con los bordes de un obstáculo opaco. La magnitud de este efecto depende fuertemente de la longitud de onda de la luz. La magnitud suele ser mínima cuando el obstáculo es un objeto macroscópico, pero si sus dimensiones son comparables con la longitud de onda de la luz, el efecto es relevante.

La difracción es causa importante del color observado en la dispersión de la luz por partículas de pequeño tamaño. Lord Rayleigh demostró que la intensidad de la luz difundida era inversamente proporcional a la cuarta potencia de la longitud de onda; en consecuencia, la luz azul es dispersada cuatro veces más que la luz roja. La disparidad en la eficiencia de la dispersión se pone de manifiesto con claridad al observar el cielo durante el día. Las moléculas, el polvo y las fluctuaciones de la densidad de la atmósfera dispersan preferentemente la luz azul y, en consecuencia, el

cielo aparece de color azul. Por otra parte, la luz directa del sol se ve privada del azul y, como resultado, aparece con una tonalidad rojiza, en especial al amanecer y al anochecer, cuando dicha luz atraviesa un espesor mayor de la atmósfera.

El mismo proceso se puede demostrar a menor escala haciendo pasar luz blanca a través de leche no homogeinizada y diluida, en la cual las partículas de grasa actúan como centros de dispersión. El líquido aparece de color azulado, en tanto que la luz transmitida directamente es rojiza. Probablemente, el color de la luna tiene su origen en un proceso similar. La dispersión producida por inclusiones mucho mayores, que no alteran el color, es la responsable del aspecto que ofrece la estrella de rubí, la estrella de zafiro y el cuarzo ojo de tigre.

La interferencia y la dispersión difractiva actúan conjuntamente en otro mecanismo generador de color: el originado en las redes de difracción. Estas redes están constituidas por un conjunto de muchas líneas o puntos igualmente espaciados entre sí; el espaciado en cuestión no debe ser demasiado grande en comparación con las longitudes de onda de la luz. La luz que atraviesa una red es dispersada en todas direcciones por cada una de las líneas que la integran y, por tanto, las ondas que proceden de dos líneas adyacentes interfieren entre sí. A una determinada longitud de onda existen ángulos de incidencia para los que la interferencia resulta ser constructiva; para cualquier otro ángulo, las ondas se eliminan una con otra. El valor real de dichos ángulos depende tan sólo de la longitud de onda y del espaciado entre las líneas o puntos de la red. Cuando la luz blanca atraviesa una red de difracción o se refleja en ella, cada una de las longitudes de onda que la integran resulta reforzada para unas determinadas direcciones emergentes. De aquí que la luz que emerge en una dirección determinada sea espectralmente pura y monocromática, pero si giramos la red se ponen de manifiesto toda una serie de espectros diferentes.

Se puede construir una red de difracción grabando líneas muy finas sobre una lámina de vidrio. Ciertos materiales naturales poseen también las dimensiones y el elevado grado de orden que exigen las redes ópticas. Por ejemplo, en los materiales denominados cristales líquidos, las moléculas están apiladas con tal regularidad que actúan como una red de difracción. Dado que el espaciado entre las moléculas depende de la temperatura, el color de dichos

materiales muestra una dependencia análoga y puede ser utilizado como indicador de la temperatura. Las filas de moléculas ordenadas que poseen algunas mariposas y escarabajos actúan como redes de difracción y dan lugar a la aparición de vivos colores en estos insectos. Los colores espectrales producidos por difracción se ponen también de manifiesto al observar la luz de una farola callejera distante a través del tejido de un paraguas o al mirar la superficie de un disco de gramófono bajo un ángulo rasante.

La red de difracción natural más importante es el ópalo. Esta piedra preciosa está constituida por esferas de dióxido de silicio y agua empaquetadas en forma compacta tridimensional con un espaciado entre sí del orden de 250 nanómetros. La matriz transparente o translúcida que llena el espacio existente entre las esferas posee una composición similar a la de éstas, pero su índice de refracción es ligeramente diferente. Cuando la luz blanca atraviesa esta red de difracción tridimensional, dentro de la piedra aparecen colores puros; los colores cambian al mover la piedra o el ojo. El mecanismo de la producción de color en el ópalo no se explicó hasta el año 1964, fecha en que se determinó la estructura de dicha piedra preciosa con ayuda del microscopio electrónico. Poco tiempo después se fabricaron los primeros ópalos sintéticos.

El espectro visible

En este catálogo de colores, que sin duda alguna no está completo, he descrito más de una docena de mecanismos diferentes. Todos estos mecanismos se pueden clasificar en cinco grandes categorías: excitaciones de átomos libres y de iones, así como vibraciones moleculares, efectos del campo cristalino, transiciones entre estados de orbitales moleculares, transiciones en las bandas de energía de los sólidos y efectos susceptibles de ser interpretados en el marco de la óptica física. El hecho de que tal diversidad de fenómenos se circunscriba en una banda de longitudes de onda cuya anchura no alcanza siquiera una octava nos puede parecer fruto de una extraordinaria coincidencia, pero la observación de que esta banda tan estrecha coincide precisamente con aquella para la cual el ojo humano es sensible todavía resulta más extraordinario.

En realidad, puede que no se trate de una mera coincidencia. En esta región del espectro electromagnético suceden tantas cosas interesantes precisamente porque éstas son las longitudes de onda para las que las interacciones de la luz con los electrones empiezan a ser importantes. Las ondas de energía más baja estimulan principalmente los movimientos de los átomos y las moléculas y, por tanto, se ponen de manifiesto generalmente en forma de calor. Las radiaciones de energía mayor pueden ionizar a los átomos y alterar permanentemente la estructura de las moléculas, de forma que sus efectos parecen ser fundamentalmente destructivos. Tan sólo en la estrecha zona de transición comprendida entre ambos extremos la energía de la luz está bien sintonizada con la estructura electrónica de la materia.

TRANSICIONES ELECTRONICAS EN LOS ATOMOS E IONES LIBRES; TRANSICIONES VIBRACIONALES EN LAS MOLECULAS	EXCITACIONES ELECTRONICAS	INCANDESCENCIA, LLAMAS, ARCOS, CHISPAS, RELAMPAGOS DESCARGAS EN GASES, CIERTOS TIPOS DE LASERES.
	VIBRACIONES	COLOR AZUL-VERDE DEL AGUA PURA Y DEL HIELO.
COLORES DEL CAMPO CRISTALINO	COMPUESTOS DE METALES DE TRANSICION	TURQUESA, LA MAYORIA DE LOS PIGMENTOS, ALGUNOS LASERES, ALGUNOS FOSFOROS, ALGUNOS MATERIALES FLUORESCENTES.
	IMPUREZAS DE METALES DE TRANSICION	RUBI, ESMERALDA, ARENISCA ROJA, ALGUNOS LASERES, ALGUNAS FLUORESCENCIAS.
	CENTROS DE COLOR	AMATISTA, CUARZO AHUMADO, VIDRIO DE AMATISTA DEL DESIERTO, ALGUNAS FLUORESCENCIAS.
TRANSICIONES ENTRE ORBITALES MOLECULARES	TRANSFERENCIA DE CARGAS	ZAFIRO AZUL, MAGNETITA.
	ENLACES CONJUGADOS	COLORANTES ORGANICOS, LA MAYORIA DE LOS COLORES DE LOS ANIMALES Y PLANTAS, EL LAPISLAZULI, LUCIERNAGAS, LASERES DE COLORANTE, ALGUNAS FLUORESCENCIAS.
TRANSICIONES EN MATERIALES QUE POSEEN BANDAS DE ENERGIA	CONDUCTORES METALICOS	COBRE, PLATA, ORO, HIERRO, BRONCE.
	SEMICONDUCTORES PUROS	SILICIO, GALENA, CINABRIO, DIAMANTE.
	SEMICONDUCTORES CON IMPUREZAS	DIAMANTE AZUL, DIAMANTE AMARILLO, DIODOS EMISORES DE LUZ, LASERES DE SEMICONDUCTORES, ALGUNOS FOSFOROS.
OPTICA GEOMETRICA Y FISICA	REFRACCION DISPERSIVA	EL ARCO IRIS, LOS DESTELLOS EN LAS PIEDRAS PRECIOSAS, ABERRACION CROMATICA.
	DISPERSION	AZUL DEL CIELO, ROJO DE LOS ATARDECERES, ROCAS LUNARES, ZAFIRO ESTRELLA.
	INTERFERENCIAS	CAPA DE ACEITE SOBRE EL AGUA, RECUBRIMIENTO OPTICO EN LAS LENTES, ALGUNOS COLORES DE LOS INSECTOS.
	REDES DE DIFRACCION	OPALO, CRISTALES LIQUIDOS, ALGUNOS COLORES DE LOS INSECTOS.

CAUSAS DEL COLOR, clasificadas aquí en 14 categorías diferentes dentro de cinco grandes grupos. Todos estos mecanismos salvo uno (las vibraciones de los átomos en las moléculas) tienen su origen en la ocurrencia de ciertos cambios en el estado de los electrones que existen en la materia. Las transiciones electrónicas son las causas más importantes del color, porque la energía necesaria para excitar un electrón corresponde a la región visible.

Los Apalaches meridionales y el crecimiento de los continentes

La prospección por reflexión sísmica de esta cordillera norteamericana induce a creer que los continentes se han desarrollado mediante el amontonamiento y la imbricación de mantos relativamente delgados de material en sus márgenes

Frederick A. Cook, Larry D. Brown y Jack E. Oliver

La teoría de la tectónica de placas –según la cual se considera a la corteza terrestre como la parte superior de un conjunto de placas rígidas que interactúan entre sí– ha contribuido en gran manera a describir la evolución del fondo océanico. Pero no ha explicado muchos aspectos relativos a la formación y la estructura de los continentes. Esta situación está empezando a cambiar, merced a la aplicación intensiva de la técnica de obtención de perfiles por reflexión sísmica al estudio de la corteza profunda y del manto superior subyacente. Se trata de una técnica de prospección empleada en la industria petrolera. Basada en la reflexión de ondas acústicas que producen las discontinuidades de densidad de la roca y de velocidad del sonido a su través, la técnica ha permitido cartografiar detalles nuevos de la estructura geológica que tiene el zócalo continental.

La aplicación de la técnica en la región meridional de los Apalaches, en los Estados Unidos, revela cómo cambian los márgenes de los continentes cuando las cuencas oceánicas se cierran y los continentes chocan en zonas de subducción –profundas fosas oceánicas próximas a márgenes continentales, donde la corteza oceánica se sumerge hacia el interior del manto–. No es nueva la sospecha de que los continentes crecen por agregación de material cortical en las zonas de subducción. Los perfiles de reflexión sísmica de los Apalaches meridionales y otros estudios geológicos relacionados con aquéllos ponen de manifiesto el desarrollo real del proceso: los continentes acrecen mediante el amontonamiento y la imbricación de delgados mantos horizontales de material cortical.

Los perfiles sísmicos los realizó el Consorcio para la obtención de Perfiles de Reflexión Continentales (COCORP), grupo de geólogos y geofísicos de la universidad, la industria y el gobierno norteamericanos dirigido por Sidney Kaufman, de la Universidad de Cornell, y dos de los autores (Brown y Oliver). El COCORP tomó sus primeros datos en el condado de Hardeman, Texas, en 1975. Desde entonces se han obtenido perfiles en 11 lugares. El estudio, consistente en la investigación de las rocas cristalinas situadas bajo estratos sedimentarios a profundidades de hasta 50 kilómetros, se ha centrado en el zócalo continental. El esfuerzo ha resultado fructífero. El COCORP ha cartografiado una masa de magma (roca líquida) existente en la parte media de la corteza en la zona central de Nuevo México, ha explorado un antiguo valle de fractura ("rift valley") sepultado en Michigan y ha detectado un importante cabalgamiento y verificado su continuidad hasta gran profundidad en la corteza, en la zona sudoccidental de Wyoming. Antes de efectuarse la prospección de Wyoming, la profundidad y la disposición del cabalgamiento, que limita un importante accidente de las Montañas Rocosas, habían sido objeto de largos debates.

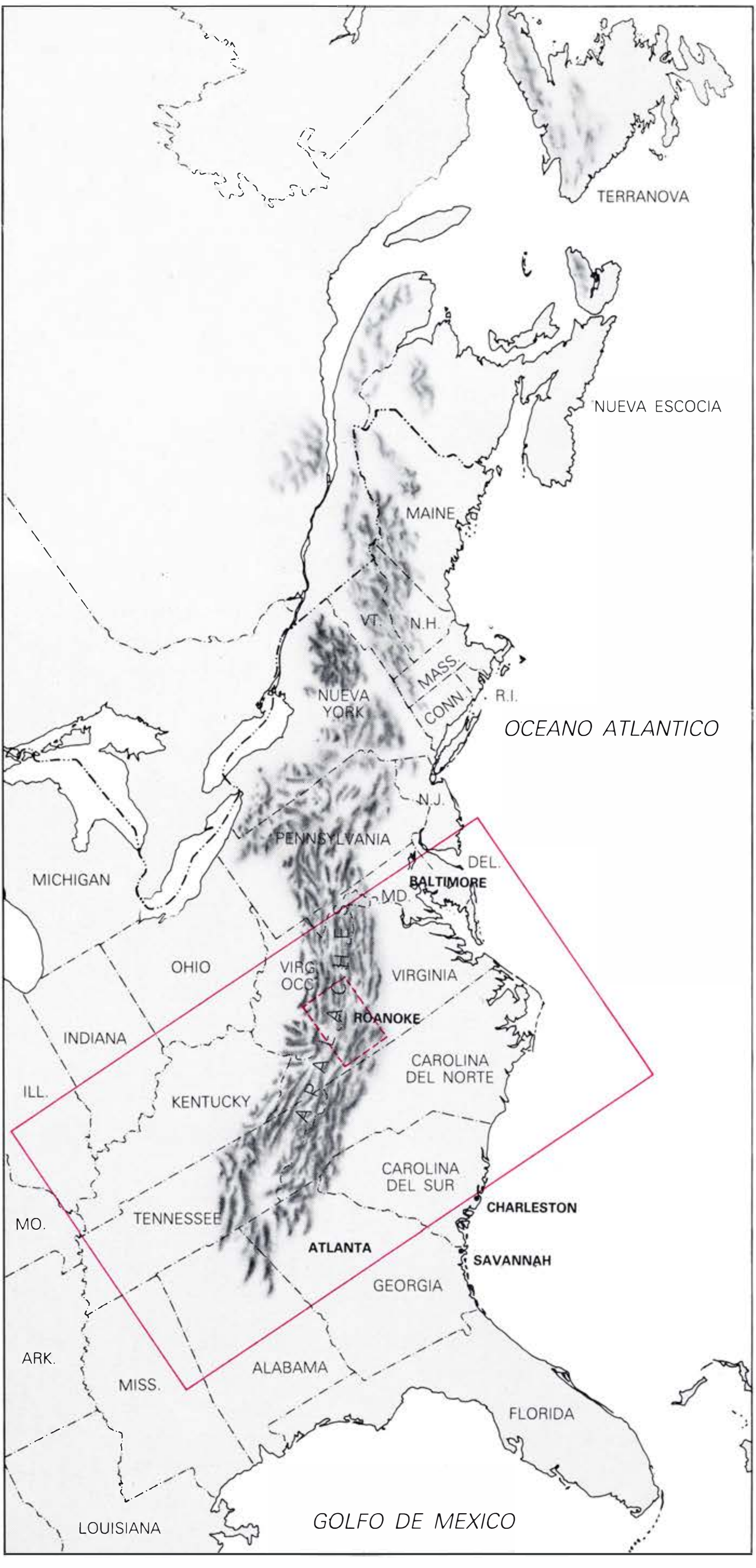
El descubrimiento más espectacular tuvo lugar en los Apalaches meridionales. Los perfiles revelaron que bajo las montañas había un paquete, de por lo menos 18 kilómetros de profundidad, constituido por capas horizontales de un material que era o había sido sedimentario. Las rocas sedimentarias constituyen uno de los tres grandes tipos de rocas; los otros dos tipos son las rocas ígneas y las metamórficas. Las rocas sedimentarias se originan a partir de sedimentos depositados en la superficie terrestre por acción del agua y del viento. Las rocas ígneas se forman por la solidificación de magma. Las rocas metamórficas se originan cuando rocas sedimentarias o ígneas quedan sometidas a la acción del calor y la presión por largos períodos de tiempo. Las rocas metamórficas pueden ser de alto o de bajo grado, según la magnitud de la recristalización que hayan experimentado en el proceso de metamorfismo.

La mayoría de las rocas superficiales de los Apalaches meridionales son rocas metamórficas muy deformadas y más antiguas que los estratos sedimentarios horizontales descubiertos bajo ellas o contemporáneas con ellos. Este hecho induce a pensar que hace unos 475 millones de años las rocas superficiales empezaron a desplazarse en forma de un manto delgado, a lo largo de por lo menos 260 kilómetros, sobre el margen continental oriental de la masa de tierra que iba a convertirse en América del Norte. Este descubrimiento confirma la hipótesis de que la fracturación horizontal o subhorizontal puede transportar enormes volúmenes de material cortical a grandes distancias y sugiere que un continente podría crecer y desarrollarse mediante el emplazamiento de delgadas láminas horizontales de material en los márgenes continentales.

Los perfiles de reflexión sísmica pueden tener una aplicación más inmediata. Han inducido a los geólogos a conjeturar sobre la existencia de yaci-

LOS APALACHES, al oeste de Roanoke, Virginia, vistos desde un satélite. La imagen, en falso color, se tomó en otoño. Roanoke es la zona verde claro de la derecha. Las zonas rojas son vegetación; las verde-azuladas brillantes, ríos y lagos. Las montañas, que alcanzan alturas de 500 a 1000 metros, se desvían aquí de una orientación nordeste a otra más hacia el norte. Las alineaciones constituyen el material sedimentario de la provincia de Valley and Ridge. La exploración por reflexión sísmica efectuada por el Consorcio para la obtención de perfiles de reflexión continentales (COCORP) cubrió parte de esta zona y terrenos de rocas cristalinas situados al sudoeste. Imagen de la Earth Satellite Corp.





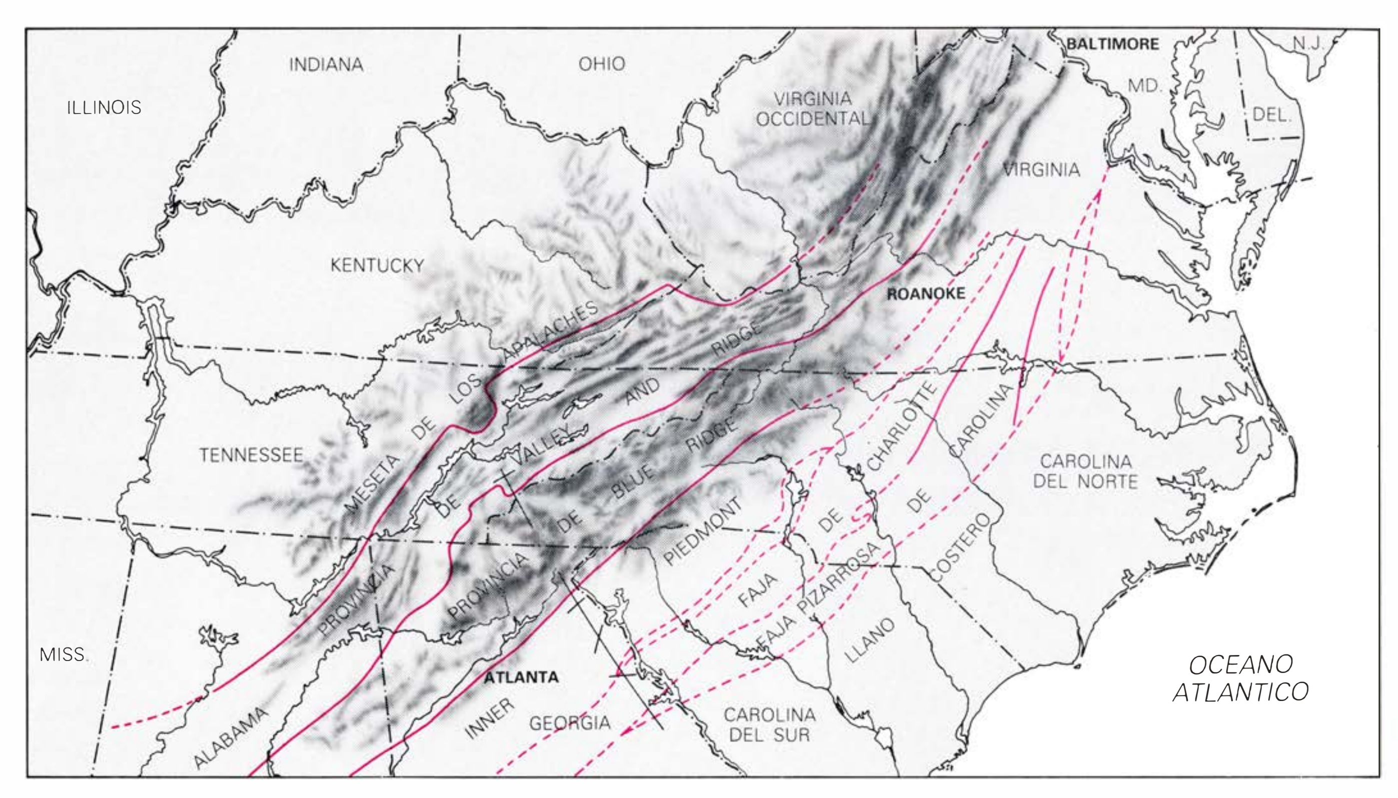
MAPA DE LA REGION ORIENTAL DE ESTADOS UNIDOS y meridional del Canadá. Puede verse la extensión de los Apalaches desde la zona central de Alabama hasta Terranova. El rectángulo interior, a trazos, señala la zona de Roanoke incluida en la imagen de satélite de la página anterior. El rectángulo exterior delimita la zona de los Apalaches meridionales representada en el mapa de la página siguiente.

mientos no descubiertos de petróleo o de gas en las rocas sedimentarias situadas bajo una parte de los mantos cabalgantes. Las rocas sedimentarias son, sin duda, las más prometedoras para la prospección de petróleo y de gas, puesto que, a diferencia de lo que ocurre con las rocas ígneas y con la mayoría de las rocas metamórficas, aquéllas no han estado sujetas a las presiones y temperaturas que hubieran podido destruir o expulsar los hidrocarburos.

Los primeros estudios de cuencas sedimentarias por reflexión sísmica se efectuaron en la década de 1920. Desde entonces, la técnica ha sido intensamente desarrollada por la industria del petróleo como parte de su esfuerzo por localizar formaciones de petróleo y gas de interés económico. La base teórica de la técnica es muy sencilla. Las ondas acústicas originadas por una explosión o por alguna otra fuente de energía acústica en la superficie se propagan radialmente hacia abajo y se ven afectadas por las discontinuidades de las capas rocosas subsuperficiales. Las ondas se reflejan allí donde hay cambios súbitos de densidad de la roca o de velocidad de las ondas. La energía reflejada se detecta en superficie por un conjunto de geófonos, es decir, de aparatos sensibles a las vibraciones.

El tiempo que tarda la onda en ir de la fuente a la discontinuidad reflectora y en volver a la superficie depende de la profundidad a la que se halla la discontinuidad. A partir de registros obtenidos por un conjunto de geófonos que respondan a ondas de trayectoria casi vertical se puede obtener información adicional respecto a la discontinuidad. Los perfiles de reflexión sísmica usados para detectar petróleo y gas responden a ondas cuyos tiempos de recorrido son inferiores a cuatro o cinco segundos, correspondientes a profundidades de ocho a diez kilómetros. Las ondas que tardan más son principalmente las que rebotan en rocas demasiado profundas como para llegar a ellas utilizando las técnicas habituales de extracción de petróleo. La prospección del COCORP se efectuó con ondas que tardaban hasta 20 segundos; estos tiempos corresponden a profundidades de 60 a 70 kilómetros.

El COCORP llevó a cabo un programa de reflexión multicanal en el que se distribuyeron hasta 2304 geófonos (dispuestos en 96 grupos de 24) a lo largo de más de 6,5 kilómetros de distancia. Toda la red de detectores se conectó por cable a un camión, donde los datos



LOS APALACHES MERIDIONALES están integrados por cuatro grandes unidades topográficas de orientación noreste-sudoeste: la provincia de Valley and Ridge, la provincia de Blue Ridge , la provincia de Piedmont (que abarca Inner Piedmont, la faja de Charlotte y la faja pizarrosa de Carolina) y el llano costero. El COCORP centró su estudio en el zócalo continental de la zona a lo largo de la línea negra. Las líneas de color indican cabalgamientos.

obtenidos se almacenaban digitalmente en una cinta magnética para su posterior procesamiento por ordenadores, del que resultaría una imagen sísmica de la geología subsuperficial. Si las fuentes acústicas y los geófonos se colocan adecuadamente, y si el procesamiento de los datos es correcto, desaparecerá el ruido de fondo indeseable y se intensificarán las señales reflejadas.

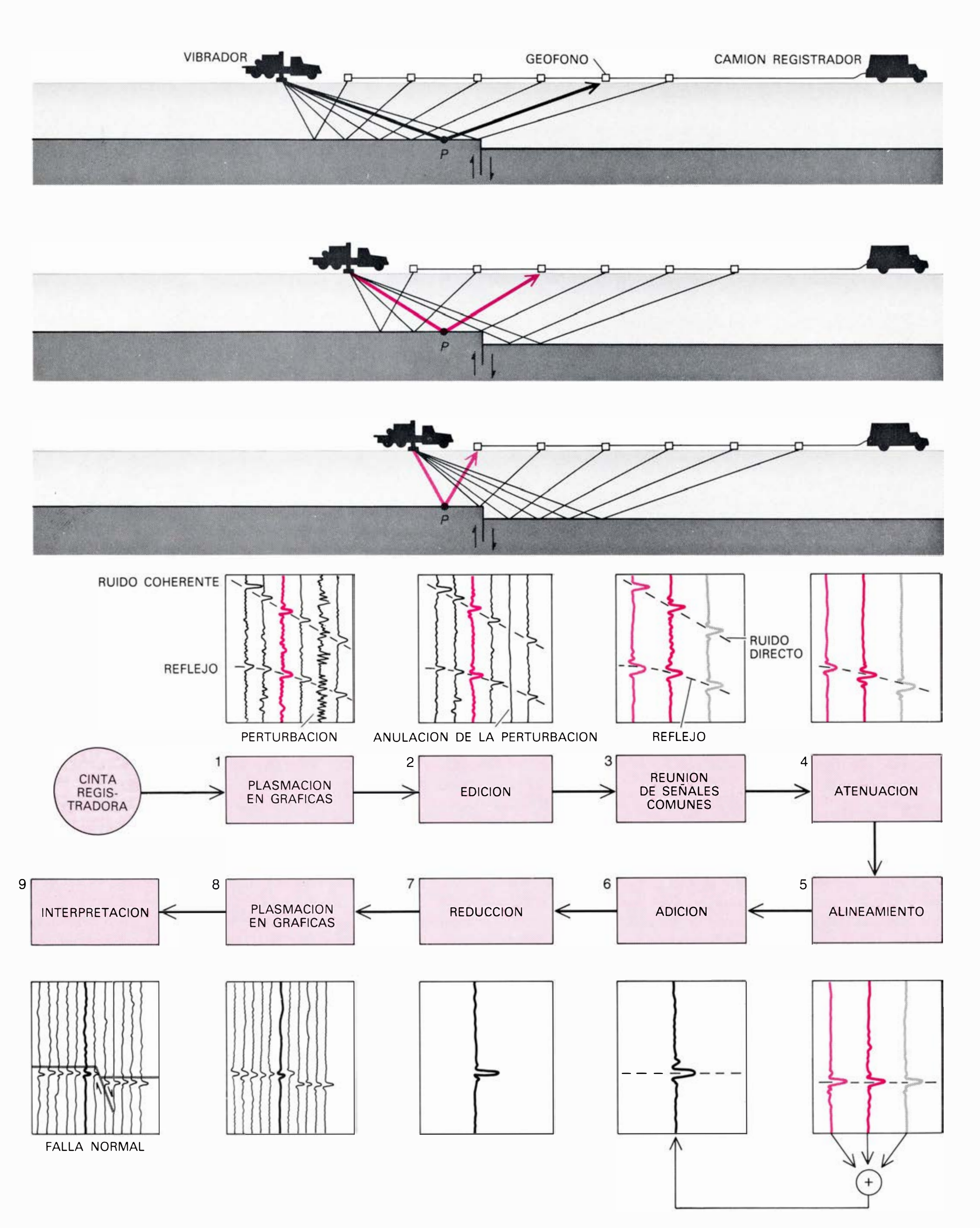
Aunque siguen empleándose explosivos como fuente de energía acústica en prospecciones petrolíferas y de gas, el COCORP y gran parte de las industrias del petróleo han optado por una fuente no explosiva y soportable por el medio ambiente, denominada Vibroseis, inventada durante la década de 1950 por la Continental Oil Company. En una prospección de COCORP típica, cuatro o cinco vibradores montados en camiones envían al subsuelo una señal de frecuencia comprendida entre ocho y 32 hertz. Durante el período de barrido de 30 segundos se varía linealmente la frecuencia para producir un "chirrido". Después de cada barrido se adelantan los vibradores unos cuantos metros y se provoca otro barrido. Cuando ya se llevan 16 barridos, que cubren una distancia de 122 metros, se repite el proceso en la siguiente estación fuente, situada a 134 metros.

Se pretende con ello registrar muchas ondas reflejadas en el mismo punto subsuperficial. La redundancia de los datos permite que en el análisis por ordenador se suprima el ruido, se intensifique la fuerza de la señal reflejada y pueda calcularse la velocidad de las ondas vibratorias subsuperficiales. Se construye luego una imagen sísmica de la estructura del zócalo continental a partir de lecturas tomadas con diferentes configuraciones de las fuentes y los receptores. El proceso recuerda la técnica médica de tomografía axial computerizada (el barrido TAC), por la cual se combinan lecturas de rayos X tomadas desde varios ángulos diferentes para obtener una representación en corte de la estructura interna del ser vivo.

La confección de perfiles por reflexión sísmica suele tardar varios meses, entre reunir y procesar los datos correspondientes a un lugar determinado. El movimiento de fuentes acústicas y detectores a través de la superficie se efectúa a un ritmo de entre uno y cuatro kilómetros por día. Reconstruir la estructura geológica subsuperficial exige el procesamiento de cantidades ingentes de datos. Por ejemplo, para analizar un trayecto noroeste-sudoeste en Georgia y Tennessee hubo que manipular unos mil millones de unidades de información: 3843 puntos de vibración, multiplicados por 96 canales, multiplicados por un tiempo de registro de 50 segundos y divididos por un ritmo de muestreo de 0,008 segundos.

El análisis por ordenador de los datos COCORP incluía la desmultiplexión (anular los múltiplos) de los registros multicanal de campo, la eliminación de datos especialmente ruidosos, el cálculo de las velocidades de las ondas, la reunión de todas las señales reflejadas por cada punto de profundidad común, la compensación relativa de las diferencias en la distancia entre fuentes y receptores, los ajustes por diferencias topográficas y de la geología subsuperficial y la superposición de las diversas señales coherentes para cada punto común de reflexión. El resultado de este procesamiento intensivo es un corte sísmico, que se parece a uno geológico, salvo en que la profundidad no está representada por distancias, sino por tiempos de trayecto, y en que las variaciones laterales de velocidad del sonido en las rocas pueden distorsionar la geometría de las líneas correspondientes a las superficies reflectoras; diferencias ambas muy importantes.

Se requiere mucha habilidad y experiencia para deducir un corte geológico preciso a partir de un corte sísmico. Por ejemplo, las interfases inclinadas están representadas en el corte sísmico por líneas que parecen desplazadas; cuanto mayor sea el buzamiento, mayor será el



PROCESAMIENTO DE LOS DATOS EN NUEVE PASOS, representado de forma muy esquemática. Los vibradores montados en camiones (*arriba*) envían al suelo una señal de baja frecuencia, que se reflejará parcialmente en las discontinuidades de la roca. La energía reflejada la detectan en superficie un conjunto de geófonos, o receptores de sonido, conectados a un camión; en él se almacenan los datos obtenidos, cuyo procesamiento ulterior revelará la geología subsuperficial. Después de cada episodio de vibración se adelantan unos metros los vibradores montados en camiones. Las líneas gruesas, negra y coloreada, indican el trayecto de una señal reflejada por un punto *P*, correspondiente a tres posiciones diferentes de un vibrador. Los datos se procesan en nueve pasos (*abajo*). Se visualizan las señales registradas (*1*) y se suprimen los ruidos (*2*). Se reúnen las señales reflejadas por un mismo punto (*3*). Se amortigua el ruido de gran amplitud (*4*) y se alinean las señales comunes (*5*), de modo que puedan sumarse en fase (*6*). Se reduce la pulsación suma (*7*) y se visualizan las señales sumadas (*8*). A continuación se interpretan los datos (*9*). El volumen de datos obliga a utilizar un ordenador.

"desplazamiento". A menudo hay que "trasladar" las líneas de reflexión para obtener una imagen precisa de la verdadera geología subsuperficial. (El perfil COCORP de los Apalaches meridionales no presentó muchos buzamientos fuertes, por lo que apenas hubo que hacer traslaciones.)

Esta es la técnica de exploración que se aplicó a los Apalaches meridionales. Los Apalaches se extienden a lo largo de más de 3000 kilómetros desde Terranova hasta la zona central de Alabama. Los Apalaches meridionales consisten en una serie de provincias geológicas distintas y de fajas de dirección noreste-sudoeste. (Provincia es un término general empleado para designar cualquier región cuyas rocas tengan una historia parecida; una faja es un accidente lineal largo cuyas rocas tienen una composición semejante.) De noroeste a sudeste, los Apalaches meridionales están formados por la provincia de Valley and Ridge, la provincia de Blue Ridge, la provincia de Piedmont (que incluye Inner Piedmont, la faja de Charlotte y la faja pizarrosa de Carolina) y el llano costero.

La provincia de Valley and Ridge se caracteriza por estratos plegados y fracturados en cabalgamientos de rocas sedimentarias formadas hace de 600 a 300 millones de años, en su gran mayoría no metamorfizadas. Los pliegues y los cabalgamientos indican que las rocas sufrieron una compresión intensa en la dirección horizontal. Durante mucho tiempo no se supo si la deformación de la provincia de Valley and Ridge afectaba también a las rocas cristalinas del zócalo subyacente a los estratos sedimentarios (en cuyo caso el proceso habría sido de tectónica de fondo) o si estaba confinada principalmente en los estratos sedimentarios superpuestos al zócalo (tectónica de cobertera). Gracias a los perfiles por reflexión sísmica y a prospecciones petrolíferas y de gas sabemos ahora que la deformación es predominantemente de cobertera. Al parecer, los estratos sedimentarios han cabalgado hacia el oeste, sobre el techo de extensas zonas horizontales de despegue por encima del zócalo cristalino, cual alfombra arrugada sobre el suelo.

Al sudoeste de la provincia de Valley and Ridge se halla la provincia de Blue Ridge. El principal accidente de transición entre ambas es un gran cabalgamiento que buza hacia el sudeste. A diferencia de las rocas sedimentarias de la provincia de Valley and Ridge, las rocas de Blue Ridge han sido, en general, muy metamorfizadas. De todos modos, las rocas de ambas provincias se deformaron por tres orogenias, es decir, tres episodios de formación de montañas: la Tacónica, la Acádica y la Alleghénica.

En el zócalo de la provincia de Blue Ridge hay rocas precámbricas cuya antigüedad se remonta, por lo menos, hasta hace mil millones de años. Muchos geólogos citaron este hecho como prueba de que la provincia de Blue Ridge está enraizada y constituye parte de la masa principal de roca del zócalo que aquí forma la corteza. Pero los nuevos datos demuestran que, si bien en esta provincia las rocas sedimentarias no metamorfizadas son raras, se hallan ocasionalmente en las "ventanas": lugares donde la erosión de las rocas metamórficas ha dejado al descubierto el material subyacente a ellas. La mayoría de las ventanas se hallan en la parte occidental de Blue Ridge, y así se ha podido saber que, al menos allí, hay rocas sedimentarias no metamorfizadas bajo las rocas cristalinas del zócalo. Es más, una de las mayores ventanas, la ventana de la montaña de Grandfather, de Carolina del Norte, se halla en la parte oriental de Blue Ridge. El afloramiento de rocas carbonatadas y argilitas cámbricas en esta zona demuestra que, al igual que en la parte occidental, también en Blue Ridge oriental hay rocas sedimentarias que han sufrido poco metamorfismo.

En el extremo oriental de la provincia de Blue Ridge encontramos un gran accidente topográfico que se extiende desde Alabama hasta Virginia. Este accidente, llamado la zona de Brevard, es una faja angosta de rocas múltiples veces deformadas que marca el límite entre las provincias de Blue Ridge y Piedmont. Se ha discutido mucho sobre la historia y la naturaleza de la deformación en la zona de Brevard. De hecho, el COCORP efectuó su primera exploración de la zona con la esperanza de conseguir nuevos datos y resolver la cuestión. Uno de los descubrimientos geológicos más importantes en la zona de Brevard lo hizo Robert D. Hatcher, Jr., de la Universidad de Carolina del Sur. Interpretó la presencia de rocas sedimentarias de un grado de metamorfismo excepcionalmente bajo como indicio de que, bajo las zonas de Brevard, debía haber estratos sedimentarios, cuyas rocas habrían sido llevadas a la superficie por fallas. Ello constituyó una nueva prueba de que el material sedimentario que se extiende bajo la provincia de Blue Ridge puede alcanzar posiciones tan orientales como la de la zona de Brevard.

Al sudeste de la zona de Brevard está Inner Piedmont; consta, principalmente, de rocas que han sufrido un metamorfismo de grado alto, atravesadas por intrusiones de rocas ígneas, como los granitos del monte Stone y de Elberton. La interpretación clásica concebía al Piedmont como el "núcleo" metamórfico de los Apalaches meridionales. Puesto que gran parte del material metamórfico parece corresponder a rocas sedimentarias que sufrieron una deformación generalizada, los geólogos creían que el Piedmont había sido levantado verticalmente, y que la deformación se había propagado hacia el oeste, hacia las provincias de Blue Ridge y de Valley and Ridge. Los nuevos datos sísmicos han modificado completamente esta concepción.

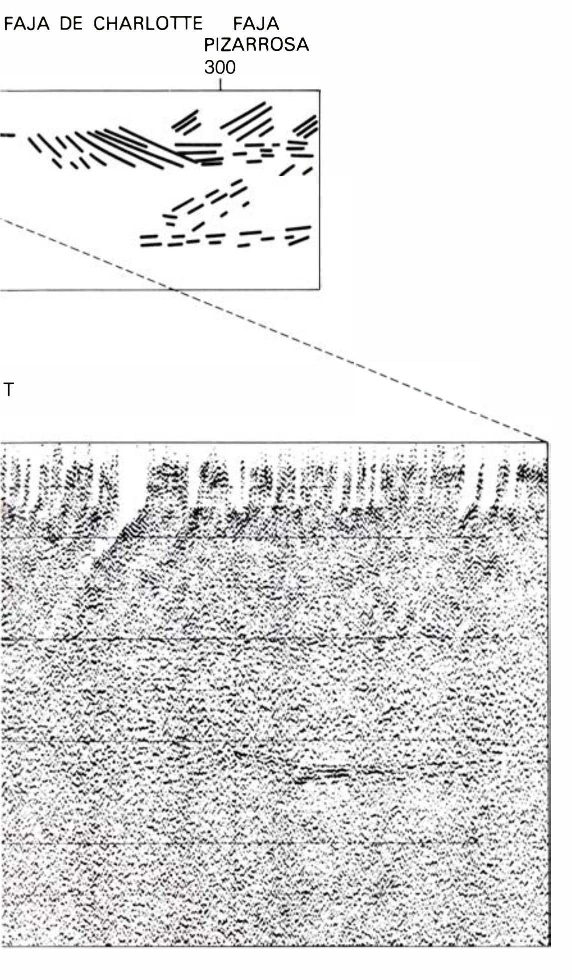
Inner Piedmont está flanqueado al sudeste por la faja del monte Kings, banda estrecha de rocas sedimentarias metamorfizadas y rocas volcánicas, algunas de las cuales pueden ser vestigios de un océano cerrado o de una cuenca geológica situada en el margen continental. Los grandes pliegues de la faja y los pliegues de la zona sudeste están alineados en dirección noreste, al igual que los pliegues de Inner Piedmont y de Blue Ridge. La faja se vio afectada, al menos, por dos períodos de deformación; uno, hace unos 450 millones de años, y, el otro, hace unos 350 millones de años.

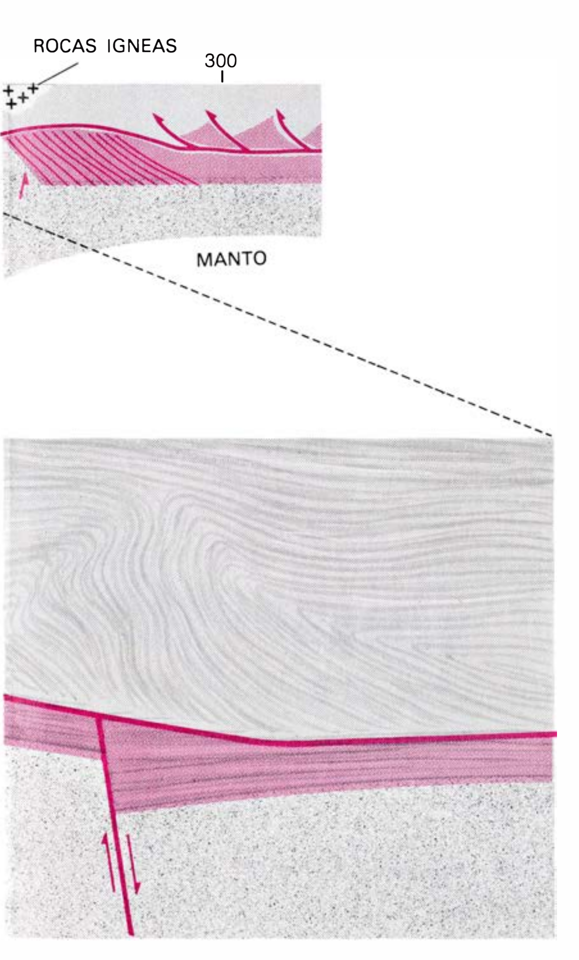
El Piedmont sudoriental está integrado por la faja Charlotte, constituida por material sedimentario metamorfizado, y por la faja pizarrosa de Carolina, formada principalmente por material volcánico metamorfizado. La faja pizarrosa de Carolina parece ser el resto de un arco volcánico como los que hay en la actualidad en el Pacífico occidental. El vulcanismo de la faja debió comenzar en las postrimerías del Precámbrico, hace de 700 a 650 millones de años, y se prolongó durante el Cámbrico hasta hace unos 500 millones de años.

El llano costero que queda al sudeste del Piedmont consta de una serie de sedimentos modernos (de menos de 200 millones de años) que cubren un zócalo cristalino. Los sondeos que penetraron en el zócalo han revelado la existencia de rocas sedimentarias metamorfizadas, como las de Inner Piedmont, y de rocas volcánicas metamorfizadas, como las de la faja pizarrosa de Carolina. Es verosímil que el zócalo sedimentario y volcánico metamorfizado de Piedmont y de la faja pizarrosa de Carolina se extienda por debajo del llano costero.

Según el cuadro histórico dibujado

VALLEY AND RIDGE
FALLA DE GREAT SMOKIES
BLUE RIDGE
FALLA DE BREVARD
INNER PIEDMONT
DISTANCIA (KILOMETROS)
0
50
100
150
200
250
TIEMPO DE IDA Y VUELTA (SEGUNDOS)
0
5
10
15
SIN DATOS
TENNESSEE
CAROLINA DEL NORTE
GEORGIA
FALLA DE BREVARD
INNER PIEDMONT
DISTANCIA (KILOMETROS)
150
175
TIEMPO DE IDA Y VUELTA (SEGUNDOS)
0
1
2
3
4
5
VALLEY AND RIDGE
FALLA DE GREAT SMOKIES
BLUE RIDGE
FALLA DE BREVARD
DISTANCIA (KILOMETROS)
50
100
150
200
PROFUNDIDAD (KILOMETROS)
0
15
30
45
ZOCALO
CRISTALINO
FALLA DE BREVARD
DISTANCIA (KILOMETROS)
INNER PIEDMONT
150
175
ROCA METAMORFICA REITERADAMENTE DEFORMADA
ROCA SEDIMENTARIA
CALIZA
ZOCALO
PROFUNDIDAD (KILOMETROS)
0
3
6
9
12
15





por la teoría de la tectónica de placas, los continentes que hoy limitan el Atlántico estaban unidos hace 200 millones de años como las piezas de un rompecabezas, formando una vasta extensión de terreno. En aquella época, América del Norte empezó a separarse de Europa, Africa y América del Sur. Puesto que los continentes estuvieron un día unidos, los datos geológicos del lado oriental del Atlántico Norte pueden facilitar la interpretación de la estructura y la formación de los Apalaches. La mayoría de los modelos concebidos de acuerdo con la teoría de la tectónica de placas para el período Carbonífero, el anterior a la apertura del Atlántico, sitúan a Africa occidental adyacente a la parte meridional de América del Norte, pero una interpretación reciente fundada en datos paleomagnéticos y debida a Edward Irving, del Departamento de Energía, Minas y Recursos del Canadá, sugiere que era la parte septentrional de América del Sur la que entonces se encontraba adyacente al este de América del Norte. En cualquier caso, Africa occidental y la parte norte de América del Sur poseen fajas de pliegues y de cabalgamientos que probablemente se originaron durante la misma orogenia del Carbonífero.

La cordillera mauritana de Africa occidental se caracteriza, de este a oeste, por una serie de fajas en cierto modo parecidas a las de los Apalaches. La parte oriental de los Mauritánides está constituida por estratos sedimentarios no metamorfizados, cubiertos en parte por rocas metamórficas que han cabalgado desde el oeste sobre los sedimentos. Al oeste hay rocas metamórficas de grado alto más antiguas, parecidas a las del Piedmont de los Apalaches meridionales. Un llano costero de rocas horizontales más modernas cubre el resto del orógeno. La cartografía geológica de la región induce a pensar que, a un período apacible de deformación, hace unos 550 millones de años, le siguió otro período de metamorfismo y fallamiento de compresión antes de la apertura del Atlántico. Probablemente, este período corresponda a la orogenia Alleghénica en los Apalaches. En un sentido amplio, los Mauritánides del Africa occidental son una imagen especular de los Apalaches.

El COCORP investigó los Apalaches en Tennessee, Carolina del Norte y Georgia. La exploración inicial cubrió una zona que se extiende desde unos 100 kilómetros al sudoeste de Knoxville (Tennessee) hasta unos 100 kilómetros al noroeste de Augusta (Georgia). El hallazgo más notorio fue una capa reflectora con buzamiento al sudeste que se extiende desde la provincia de Valley and Ridge y se hunde bajo Blue Ridge e Inner Piedmont, a una profundidad de entre cuatro y 10 kilómetros bajo la superficie.

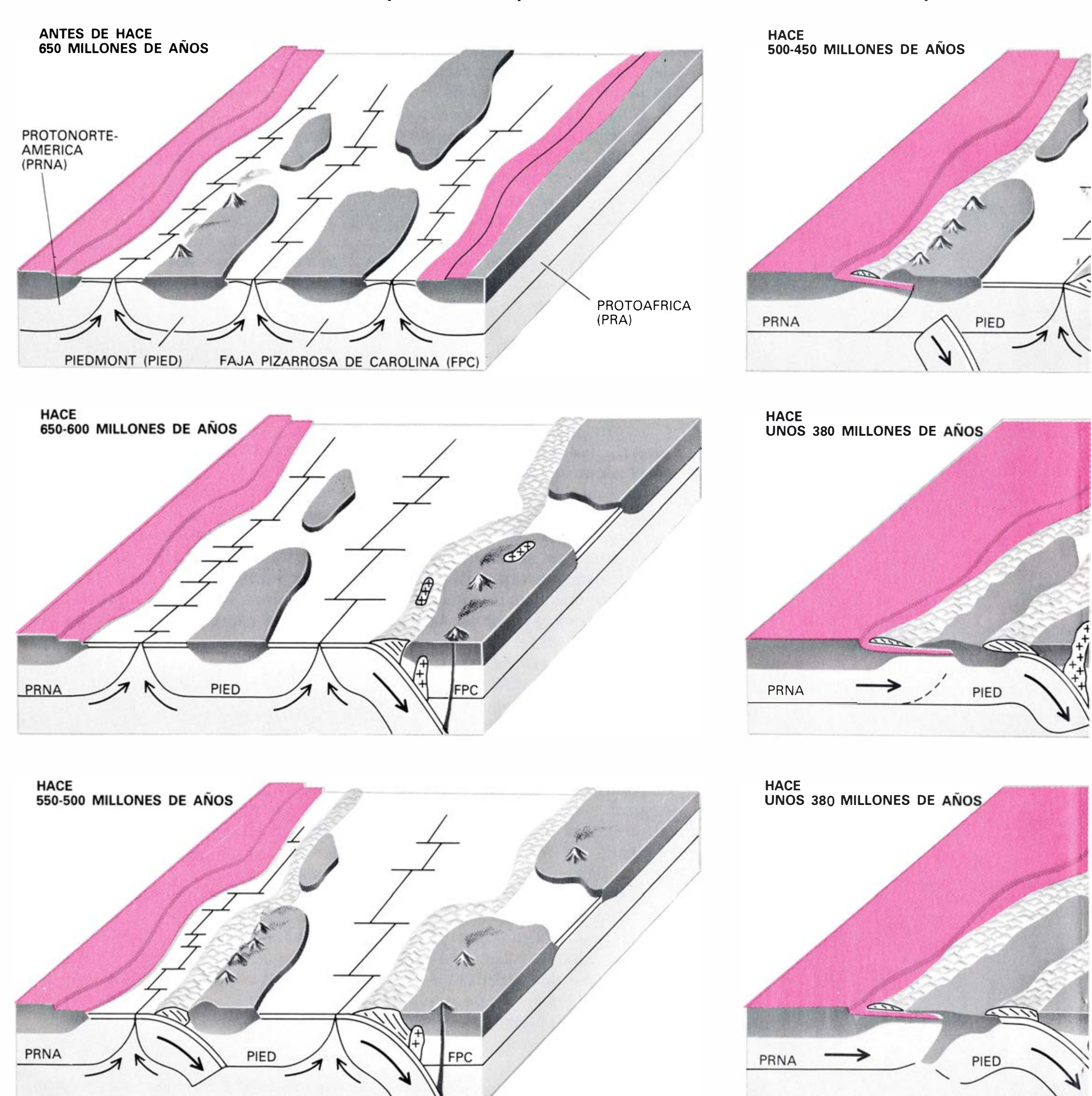
Si bien son varios los tipos y las configuraciones de rocas que podrían dar lugar a las reflexiones, interpretamos el material reflector como capas de sedimentos estratificados. La interpretación se basa en cuatro clases de pruebas. Primero, las líneas correspondientes a superficies reflectoras situadas bajo Blue Ridge e Inner Piedmont pueden seguirse hasta unidades parecidas de la provincia de Valley and Ridge, y correlacionarse con ellas, cuyas rocas son probablemente sedimentarias. De hecho, algunas partes del perfil por reflexión sísmica de Blue Ridge y del Piedmont se parecen mucho al perfil por reflexión sísmica de Valley and Ridge. Segundo, la presencia de rocas sedimentarias en las ventanas de Blue Ridge indica que las rocas cristalinas de esta zona recubren material sedimentario. Tercero, el descubrimiento de rocas carbonatadas raras en la zona de Brevard sugiere que éstas fueron arrancadas de capas sedimentarias subyacentes por actividad a lo largo de los labios de la falla. Por último, los resultados del COCORP se parecen a los de las exploraciones por reflexión sísmica de los márgenes continentales actuales, que, evidentemente, están constituidos por material sedimentario.

H. Clark y sus colaboradores, del Instituto Politécnico de Virginia, han realizado estudios locales por reflexión sísmica de la zona situada 100 kilómetros al norte de la región explorada por el COCORP en Carolina del Norte. Los estudios indican que también bajo la zona de Blue Ridge hay capas

PERFIL DE REFLEXION SISMICA de los Apalaches meridionales, representado arriba esquemáticamente. El eje vertical indica los tiempos de ida y vuelta de las ondas reflejadas. El eje horizontal señala la posición de los principales accidentes geológicos e indica la distancia, a lo largo de la superficie, desde el extremo noroccidental de la zona donde empieza la exploración del COCORP. También se aprecian las fronteras entre estados. Debajo del perfil esquemático se ofrece una fotografía del perfil real perpendicular a la falla de Brevard, faja estrecha de rocas reiteradamente deformadas que señala el límite entre las provincias de Blue Ridge y Piedmont. Abajo, una posible interpretación geológica de los perfiles. En gris, rocas cristalinas de origen predominantemente continental. Las líneas grises onduladas representan rocas metamórficas plegadas próximas a la superficie. Los colores claros representan estratos sedimentarios. Las fallas son las líneas con puntas de flecha que indican el movimiento relativo.

estratificadas horizontales. Por consiguiente, es probable que los estratos se extiendan por debajo de gran parte de los Apalaches meridionales. Leonard Harris y Ken Bayer, del Servicio Geológico de Estados Unidos, han analizado datos obtenidos recientemente por reflexión sísmica en Carolina del Norte; esos datos demuestran igualmente la existencia de capas estratificadas que se extienden desde la provincia de Valley and Ridge bajo Blue Ridge hasta el Piedmont.

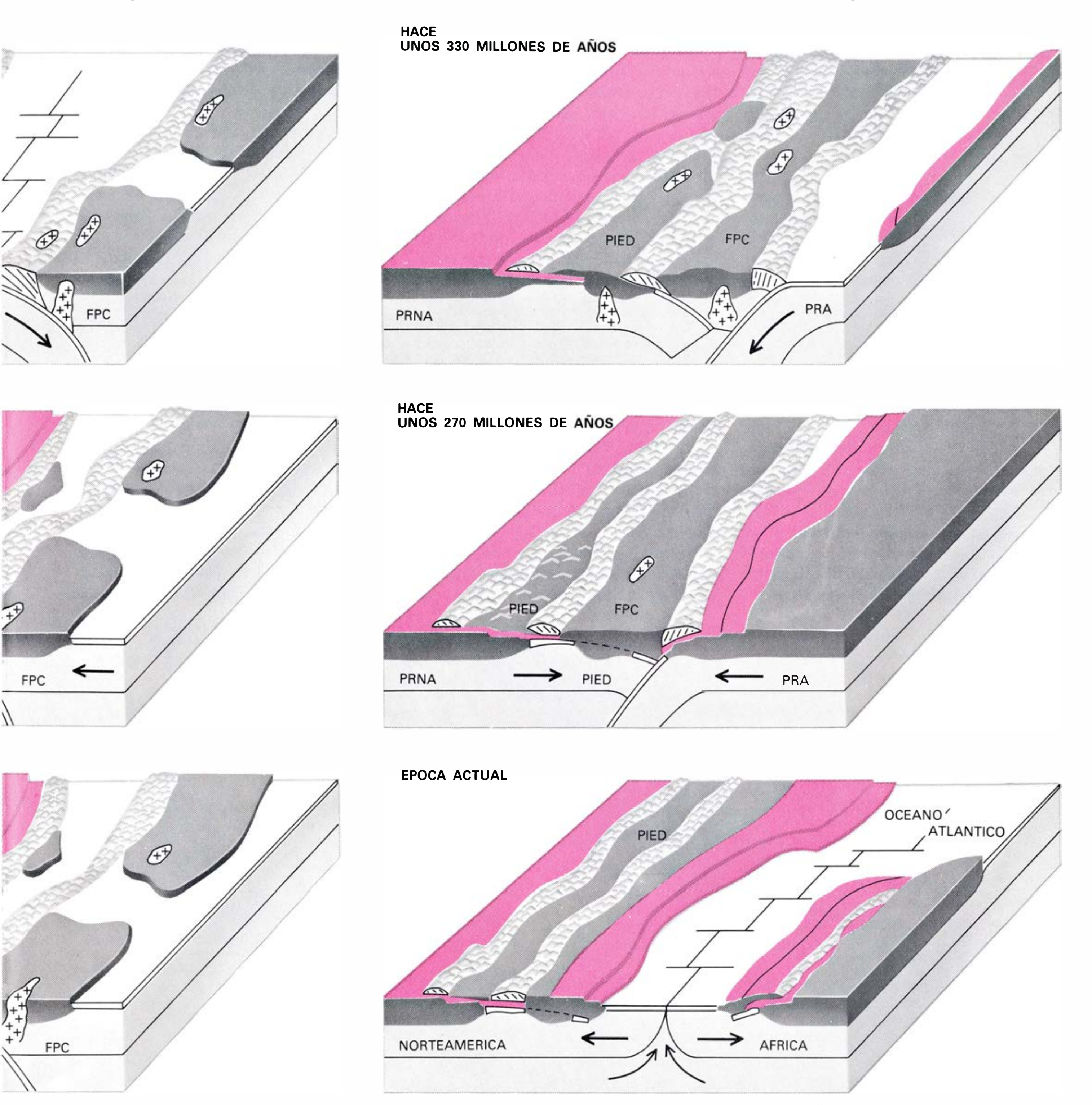
La única explicación que cabe dar a la presencia de los estratos sepultados en profundidad es que las rocas cristalinas que los cubren se emplazaron a lo largo de un importante cabalgamiento subhorizontal (falla horizontal subsuperficial). Los datos del COCORP indican también que la falla de Brevard es un cabalgamiento ramificado o desprendido del gran cabalgamiento subhorizontal. La ramificación pudo haber



HISTORIA TECTONICA de los Apalaches meridionales representada en una serie de nueve bloques diagrama. El gris oscuro significa corteza continental; el blanco, corteza oceánica, y el color claro, rocas sedimentarias de la plataforma continental. La parte sólida superficial de la tierra, que alcanza un grosor de hasta 100 kilómetros, es un conjunto de placas rígidas con interacción mutua que pueden ser absorbidas en el interior del manto en zonas de subducción: fosas oceánicas profundas donde dos placas chocan y una se sumerge, introduciéndose bajo la otra. El choque de un continente o de un arco de islas oceánicas con un complejo de zona de subducción provoca el combamiento de las rocas de superficie. En este caso, el plegamiento dio lugar a tres orogenias, es decir, a tres episodios de formación de montañas, en los Apalaches meridionales. Hace unos 650 millones de años, durante el Precámbrico (*1*) una extensión megacontinental se dividió en dos placas continentales (Laurencia, o Protonorteamérica, y Gondwana, o Protoáfrica y Protosudamérica) y varios fragmentos menores que comprendían el Piedmont y la faja pizarrosa de Carolina. La subducción provocó vulcanismo en

originado también otras fallas de la zona.

En el estudio del COCORP se detecta la extensión continua de los reflectores estratificados horizontalmente hasta debajo de Inner Piedmont y la faja de Charlotte. Dado que las líneas correspondientes a estos reflectores son parecidas a las que subyacen a Blue Ridge y tienen, además, continuidad lateral con ellas, creemos que, en esencia, las capas sedimentarias son continuas desde la provincia de Valley and Ridge hasta la faja de Charlotte. Con todo, hay un cambio sustancial en el carácter de los reflectores a una distancia de unos 250 a 300 kilómetros del extremo noroccidental del perfil. Una serie de reflectores que buzan hacia el este, parecida a la que suele caracterizar a los sedimentos de un talud continental, diferencia esta parte del corte sísmico de la banda de reflectores de la parte occidental.



la faja pizarrosa de Carolina, probablemente hace unos 650 millones de años (*2*). La actividad volcánica persistió hasta hace unos 500 millones de años (*4*). El cierre de la cuenca comprendida entre Protonorteamérica y el Piedmont (*3, 4*) pudo haber sido el causante de la orogenia Tacónica, ocurrida hace de 500 a 450 millones de años. El cabalgamiento de cobertera que se inició en esa época pudo haber sido provocado por dos procesos. El primer proceso (*4, 5*) implica la subducción continuada hacia el este de la parte profunda de Piedmont y el descortezamiento de los 10 a 15 kilómetros superficiales. El otro proceso (*5'*) conlleva la subducción hacia el oeste de la parte profunda de Inner Piedmont y de Blue Ridge, mientras la parte superior empieza a constituir el manto cabalgante. El cierre de la cuenca oceánica comprendida entre Inner Piedmont y la faja pizarrosa de Carolina (5, 6) dio lugar a la orogenia Acádica, hace de 400 a 350 millones de años, y al cabalgamiento continuado de Piedmont y del continente norteamericano. La orogenia Allegheniana corresponde al choque entre Protoáfrica y Norteamérica (*6, 7*). La apertura del Atlántico ocurrió hace unos 200 millones de años (*8*).

En superficie, sobre el reflector inclinado, se halla el granito de Elberton, una gran masa de roca ígnea que hizo intrusión en la corteza. Bajo las rocas plegadas y metamorfizadas de la parte oriental de la faja de Charlotte hay otra serie discreta y multiestratificada de reflectores horizontales. La cuña de reflectores inclinados, tanto bajo el granito de Elberton como bajo la faja de Charlotte, junto con los reflectores horizontales subyacentes a la faja, sugieren la presencia de una potente serie estratificada de rocas a una profundidad de 12 a 18 kilómetros bajo la superficie.

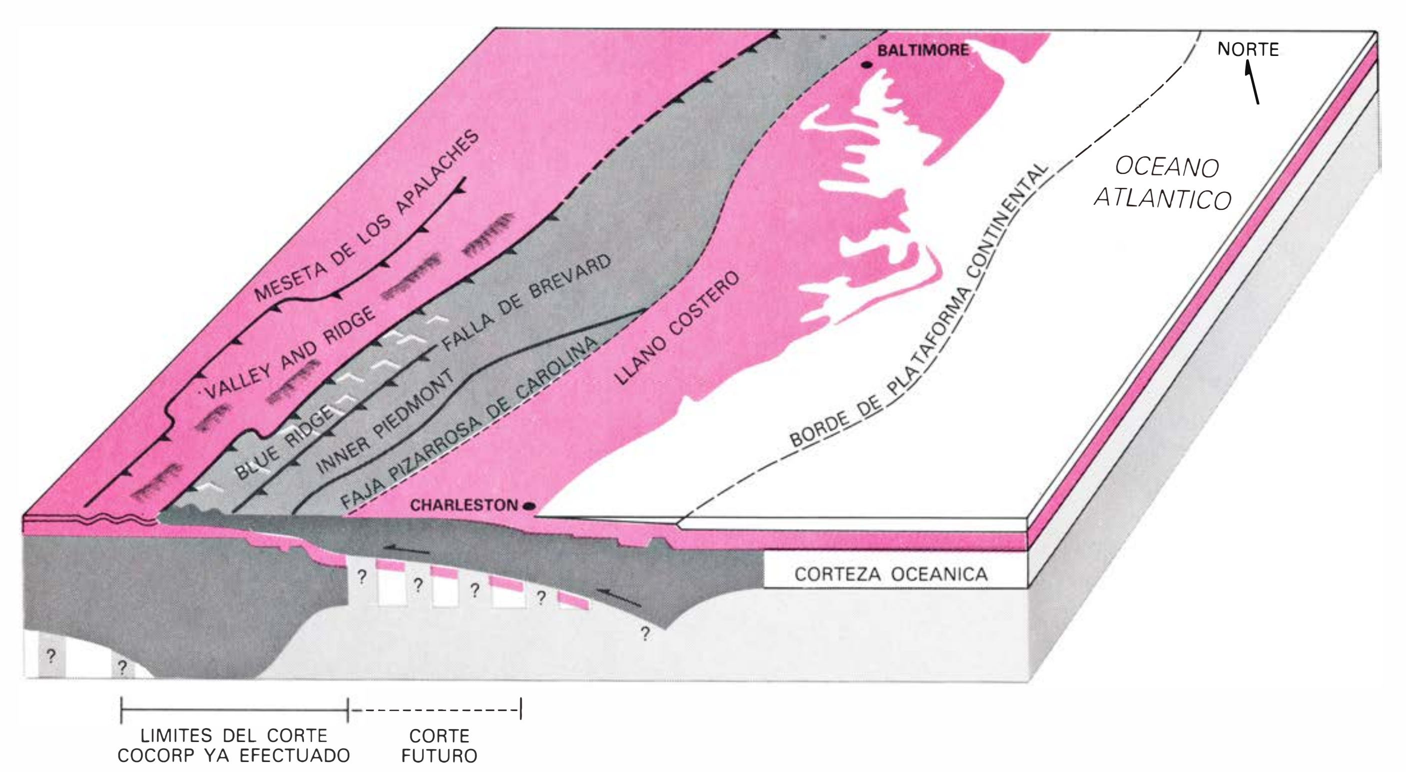
En el extremo sudoriental de la zona explorada se observan otros cambios significativos en la respuesta acústica de la parte profunda de la corteza. Hay reflectores horizontales anómalos de 10,5 a 11 segundos (equivalentes a profundidades de 30 a 33 kilómetros) que podrían corresponder a la transición entre la corteza y el manto. Esta transición se conoce como discontinuidad de Mohorovičić. En los dos tercios noroccidentales del perfil no se observan líneas de reflexión correspondientes a la discontinuidad de Mohorovičić ni tampoco a muchos otros niveles reflectores profundos. Otros datos sísmicos inducen a creer que bajo Inner Piedmont y Blue Ridge, la corteza puede llegar a profundidades de entre 40 y 45 kilómetros. En el extremo sur, sobre la discontinuidad de Mohorovičić hay reflectores intrigantes que buzan hacia el oeste, cuya interpretación resultará difícil en tanto no se extienda el perfil hacia el este. Podrían representar zonas de falla, partes de un antiguo complejo de zona de subducción o, tal vez, hasta una estructura geológica más antigua no relacionada con los Apalaches.

No se sabe todavía qué representan los cambios de reflexión en la transición entre Inner Piedmont y la faja de Charlotte. Para unos, los reflectores que buzan al este bajo la faja de Charlotte señalan los puntos donde el cabalgamiento principal se verticaliza y se sumerge hacia zonas profundas de la corteza. Pero no es fácil explicar, en el contexto de esta hipótesis, los reflectores horizontales del sudeste subyacentes a la faja pizarrosa de Carolina; por eso optamos por otra interpretación. Creemos que los reflectores horizontales y la gruesa serie de reflectores que buzan hacia el este bajo la faja de Charlotte son rocas sedimentarias que formaban parte de un margen continental y de una cuña de plataforma continental, hoy sepultadas bajo el manto cabalgante. Estas rocas sedimentarias se hallan a gran profundidad, y, probablemente, han sido metamorfizadas.

Nuestra hipótesis permite explicar además la disminución de profundidad de la discontinuidad de Mohorovičić desde el noroeste hasta el sudeste de la zona explorada: es consecuencia de una transición desde la corteza continental, al oeste, hasta una antigua corteza oceánica o una corteza continental adelgazada, al este. Ello incita a pensar que las rocas deformadas de la faja de Charlotte y de la faja pizarrosa de Carolina, al igual que las rocas deformadas de Blue Ridge y de Inner Piedmont, cabalgaron sobre un paquete de sedimentos. De acuerdo con esta interpretación, el cabalgamiento debe proseguir hacia el este. Los estudios sísmicos que se están llevando a cabo deberán dilucidar si ello es o no así.

Si los reflectores de la faja de Charlotte y de la faja pizarrosa de Carolina son realmente capas sedimentarias o de sedimentos metamorfizados, cabe deducir que, antes de que empezara el cabalgamiento, las rocas muy metamorfizadas y deformadas de Blue Ridge debían hallarse sumamente lejos de su posición actual con respecto al interior continental. Probablemente, se encontraban al este de la posición que hoy ocupa la faja pizarrosa de Carolina. La distancia desde el extremo sudoriental de la región explorada hasta el extremo occidental de Blue Ridge, unos 260 kilómetros, es la distancia lateral mínima de desplazamiento de las rocas de esta última sierra.

¿Qué revelan los perfiles de reflexión sísmica acerca de la historia tectónica de los Apalaches? Según la teoría



GEOLOGIA SUBSUPERFICIAL DEL ESTE de Estados Unidos, deducida en base a los perfiles de reflexión sísmica de la prospección COCORP. Se indica una ampliación programada de la zona de exploración. El color representa los estratos sedimentarios, el gris oscuro corresponde a rocas cristalinas de origen predominantemente continental y, el blanco, a la corteza oceánica basáltica. Hace de 500 a 250 millones de años cabalgó sobre la plataforma continental un manto de roca cristalina. Durante 200 millones de años se ha estado formando una nueva plataforma frente a la costa atlántica.

de la tectónica de placas, las placas rígidas de la litosfera, la parte superficial sólida de la tierra que alcanza una profundidad de unos 100 kilómetros, flotan sobre la astenosfera, la capa móvil y moderadamente fluida del manto cuyo espesor es de unos centenares de kilómetros. Cada placa consta de una capa superior de corteza (cuyo grosor mide de 35 a 40 kilómetros en las placas continentales y de 5 a 10 kilómetros en las oceánicas) y una capa inferior de manto sólido y rígido. La litosfera oceánica se genera en las dorsales mesoceánicas, donde continuamente está manando magma que se enfría y endurece para formar el borde remolcador de la placa en movimiento. Esta litosfera se reabsorbe finalmente en el interior del manto en las zonas de subducción, profundas fosas oceánicas formadas allí donde dos placas chocan y una se sumerge bajo la otra.

Una zona de subducción consume litosfera hasta que se ve afectada por un continente o un arco de islas. Puesto que el material de la corteza continental es mucho más ligero que el del manto, la mayoría de los geólogos consideran que un continente difícilmente puede subducir en su totalidad. Si realmente no puede, entonces la colisión de un continente o de un arco de islas con una zona de subducción provocará un cambio drástico en la naturaleza de la subducción e incluso la paralizará. La deformación de la litosfera en el lugar de la colisión puede originar una cordillera.

Los Apalaches meridionales se han desarrollado por una serie de choques de fragmentos de material continental o de arcos de islas en el borde oriental de América del Norte durante las orogenias Tacónica, Acádica y Alleghénica. Se habían propuesto varios modelos de procesos evolutivos, pero los datos del COCORP les imponen nuevas limitaciones. Analizaremos un modelo prometedor que desarrolló Hatcher para hacer encajar los datos de la geología de superficie. Hemos modificado el modelo para que integre la información subsuperficial del COCORP. Las fechas de los episodios orogénicos se han determinado en base a la deposición de sedimentos en la cuenca apalache y a la datación radiactiva de rocas ígneas y metamórficas. La composición global que debían tener las masas que chocaron se ha deducido a partir de los tipos de roca presentes en las distintas provincias y fajas.

Hace unos 750 millones de años, el magma que ascendía del interior de la tierra dividió una extensión megacontinental en, por lo menos, dos vastos continentes (Laurencia, o Protonorteamérica, y Gondwana, o Protoáfrica) y, por lo menos, dos fragmentos continentales, que comprendían el fragmento Inner Piedmont- Blue Ridge y el fragmento de la faja pizarrosa de Carolina. En el período de fragmentación inicial, las rocas que hoy son el material volcánico y sedimentario metamorfizado de Blue Ridge se depositaron en la cuenca que separaba Protonorteamérica del fragmento Inner Piedmont-Blue Ridge.

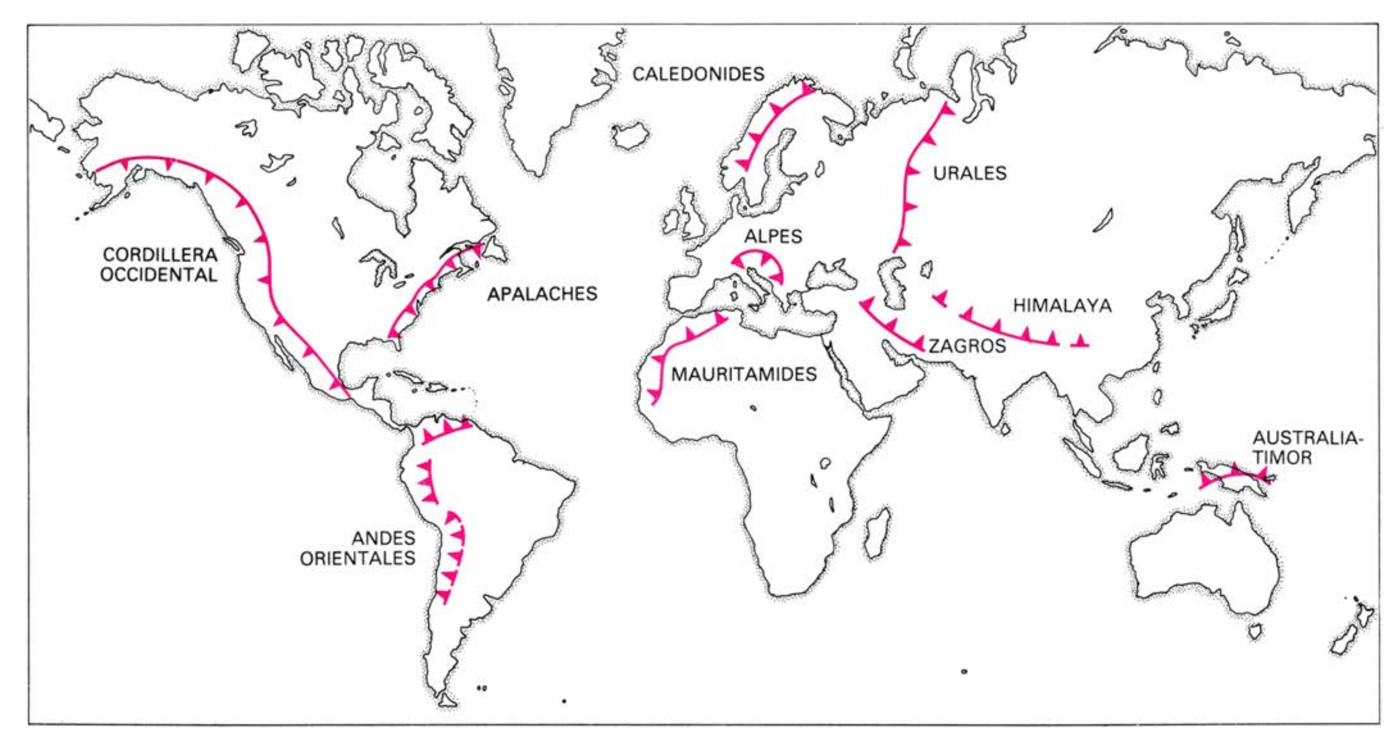
No hay pruebas irrefutables de que Piedmont y la faja pizarrosa de Carolina procedieran del mismo megacontinente del que derivó América del Norte. Sin embargo, el hecho de que algunas rocas de Blue Ridge y Piedmont tengan la misma edad según datación radiactiva (unos 1000 millones de años) que las rocas del zócalo del este de Norteamérica induce a creer que estos trozos de corteza experimentaron un proceso de metamorfismo más o menos contemporáneo, durante el Precámbrico. Por tanto, pudieron formar parte del mismo bloque continental. Lo más verosímil es que la estructura de los Apalaches no sea consecuencia del choque de un único bloque cortical con América del Norte, sino de colisiones de pequeños fragmentos continentales o de arcos de islas, colisiones que serían análogas a los numerosos choques que se producen hoy en el Pacífico sudoccidental, donde Australia se mueve hacia Asia con un grupo de arcos de islas y de fragmentos continentales atrapados entre ellas.

En el arco de islas que constituía el fragmento correspondiente a la faja pizarrosa de Carolina, el vulcanismo empezó hace unos 650 millones de años. Ello significa que la subducción, que dio lugar a la actividad volcánica, debió también comenzar en esa época. Hace unos 500 millones de años, la cuenca comprendida entre Protonorteamérica y el fragmento Inner Piedmont-Blue Ridge empezó a cerrarse a consecuencia de la subducción. La existencia de capas sedimentarias bajo Blue Ridge y Piedmont implica que la zona de subducción buzaba hacia el este; una subducción hacia el oeste en aquel período probablemente habría provocado la destrucción de los sedimentos durante el vulcanismo. Con anterioridad a una fecha que podríamos datar hace entre 450-500 millones de años los sedimentos que formaron las areniscas y las argilitas que hoy se encuentran en la provincia de Valley and Ridge procedían de Protonorteamérica. En los 250 millones de años siguientes los sedimentos procedieron principalmente del lado opuesto (del este). Este cambio en la historia sedimentaria coincide con el primer episodio de deformación y metamorfismo (la orogenia Tacónica, hace entre 500 y 450 millones de años) y puede atribuirse al resultado del cierre de la cuenca comprendida entre el fragmento Inner Piedmont-Blue Ridge y Protonorteamérica y el posterior choque de ambas masas de tierra. Las rocas sedimentarias del este procedieron del manto cabalgante que empezó a moverse hacia el oeste sobre el continente.

No está claro cómo un manto de 10 a 20 kilómetros de grosor se despegó de la corteza inferior de Inner Piedmont y de Blue Ridge y cabalgó luego sobre la plataforma continental. ¿Por qué se rompió por donde lo hizo y qué ocurrió con los 80 kilómetros restantes de litosfera subyacente? Hay una hipótesis que afirma que, cuando la corteza superior se despegó, la corteza inferior y el manto superior restantes prosiguieron la subducción hacia el este y chocaron con el arco de islas de la faja pizarrosa de Carolina. Para otra hipótesis parecida, la polaridad de la subducción en la cuenca se invirtió, de suerte que una "escama" de la corteza superior se vio empujada hacia el continente, mientras que la corteza inferior de Inner Piedmont y de Blue Ridge fue subducida hacia el oeste. (La hipótesis de la escamación había sido adoptada por E. Oxburgh para explicar algunos accidentes de los Alpes.)

El segundo episodio de formación de montañas fue la orogenia Acádica, hace de 400 a 350 millones de años. Caracterizada por metamorfismo y deformación generalizados, la orogenia se desencadenó por el cierre de la cuenca oceánica comprendida entre el fragmento Inner Piedmont-Blue Ridge y el fragmento de la faja pizarrosa de Carolina. Por aquel entonces, el fragmento Inner Piedmont-Blue Ridge probablemente se agregó al protocontinente norteamericano por cabalgamiento. La actual faja del monte Kings podría ser el vestigio superficial de la antigua zona de colisión entre el arco de islas de la faja pizarrosa de Carolina e Inner Piedmont. Se desconoce qué tamaño tuvo el fragmento de la faja pizarrosa de Carolina, pero éste pudo ser muy extenso, ya que bajo el llano costero se extienden rocas metamórficas que se parecen mucho a las rocas de la faja pizarrosa de Carolina.

Después de la orogenia Acádica, el siguiente (y último) acontecimiento compresivo importante de los Apala-



ALGUNAS CORDILLERAS DEL MUNDO (la cordillera occidental americana, los Andes orientales, los Mauritánides, los Alpes, los Urales, los Caledónides, el Zagros, el Himalaya y Australia-Timor) pueden consistir, lo mismo que la cordillera de los Apalaches, en mantos cabalgantes delgados.

ches meridionales fue la orogenia Alleghénica, hace de 300 a 250 millones de años. Este episodio de formación de montañas puede atribuirse al choque de Protonorteamérica con Protoáfrica (o, quizás, con Sudamérica) para formar el supercontinente de Pangea. Si bien la orogenia Alleghénica no fue tan destacada en los Apalaches septentrionales como lo fueron las otras dos, en los meridionales dio lugar a cabalgamientos de gran escala y a una actividad ígnea generalizada.

Por esa época se quebró la superficie en la zona de Brevard, permitiendo el transporte desde abajo de rocas sedimentarias carbonatadas. La datación radiactiva indica que muchas de las masas ígneas de Piedmont se emplazaron hace entre 300 y 250 millones de años. Cabalgamientos a gran escala producidos durante este mismo período se han cartografiado también en Africa occidental, si bien todavía no se ha determinado su límite occidental. En nuestra opinión, un segmento de la plataforma continental africana (o sudamericana) debió introducirse bajo el margen oriental del fragmento de faja pizarrosa de Carolina, y se formó una faja de pliegues y cabalgamientos vergentes en sentido opuesto.

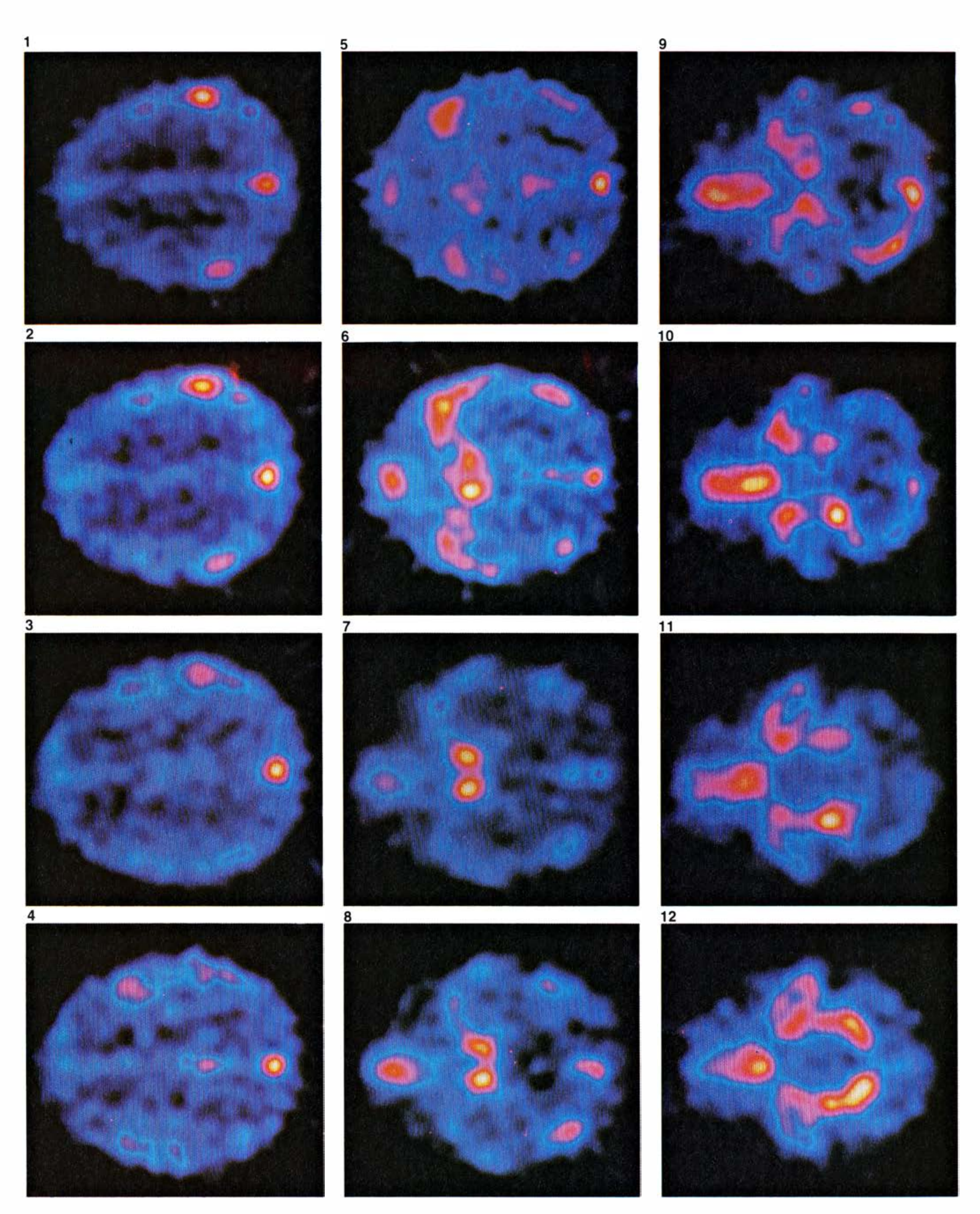
Después de la orogenia Alleghénica, nuevos episodios de tectónica de extensión prosiguieron la desintegración de un megacontinente (el Pangea) en continentes menores; de eso hace entre 250 y 200 millones de años. A medida que los continentes se distanciaron, y a consecuencia de ello, se originó el océano Atlántico. Las cuencas triásicas de la costa oriental estadounidense, como las fosas tectónicas (cubetas de laderas casi verticales) de New Jersey y Connecticut, se formaron en esa época. Con el crecimiento del Atlántico se fue construyendo la actual plataforma continental frente a la costa oriental (y frente a la occidental de Africa y a la septentrional de Sudamérica).

¿Podemos aventurar que los mantos de corrimiento delgados y de gran escala sean un resultado general de choques continentales? Hay varios indicios que apuntan en ese sentido. Hemos examinado las fajas de pliegues y cabalgamientos de Africa occidental y de Sudamérica septentrional, que tienen una estructura geológica muy parecida a la de las fajas apalaches. David Gee, del Servicio Geológico de Suecia, ha cartografiado grandes cabalgamientos horizontales en las montañas caledonianas de Escandinavia, algunos de los cuales aparentan desplazamientos de centenares de kilómetros.

El tipo de cabalgamiento de cobertera hallado en la provincia de Valley and Ridge puede verse también en las fajas de pliegues y cabalgamientos de la cordillera de Montana y Alberta, que forma parte de las Rocosas. En esta zona, el COCORP tiene previsto realizar estudios para determinar el límite occidental del cabalgamiento. Los cabalgamientos de cobertera pudieron haber predominado también en otras cordilleras, como los Alpes, el Himalaya y el Zagros del Oriente Medio. Algunos geólogos creen que el tipo de cabalgamiento apreciable en los Apalaches está produciéndose hoy en la región de Timor del Pacífico sudoccidental, donde la plataforma continental australiana se ha introducido unos 150 km bajo el arco de islas de Timor.

El aprisionamiento de rocas sedimentarias bajo extensos mantos metamórficos tiene consecuencias importantes para la obtención de gas y petróleo. El descubrimiento de la existencia de material sedimentario bajo las rocas metamórficas de Blue Ridge y Piedmont exige considerar la posibilidad de extraer hidrocarburos de esta parte de los Apalaches. Puede que el metamorfismo de los sedimentos no haya sido lo suficientemente intenso como para eliminar todos los yacimientos de hidrocarburos de las rocas situadas bajo el manto cabalgante.

El descubrimiento de depósitos sedimentarios bajo el núcleo de los Apalaches tiene implicaciones que rebasan el ámbito de la cordillera. El mecanismo de cabalgamiento de cobertera con participación de rocas del zócalo pudo haber sido responsable de la formación de sistemas montañosos desde que empezaron a actuar procesos de tectónica de placas. En este caso, la evolución de los continentes se caracterizaría, no por el adosamiento lateral a lo largo de bordes verticales, sino por la imbricación y el amontonamiento de mantos de corrimiento.



VOLUMEN DE SANGRE en el cerebro de un hombre normal en imágenes, que corresponden a 12 niveles craneanos diferentes, obtenidas mediante tomografía de emisión de positrones (TEP). El dibujo de la página opuesta define los niveles correspondientes a las imágenes. Las regiones coloreadas de la figura indican la distribución, en cráneo y cerebro, del trazador radiactivo carbono 11 (^{11}C), que tiene una semivida de tan sólo 20 minutos. El paciente inhala el trazador en forma de ^{11}C-monóxido de carbono; éste se combina con la hemoglobina, en los glóbulos rojos, y es trasportado rápidamente a través del cuerpo. Al desintegrarse, un átomo de carbono 11 libera un positrón, o electrón positivo, de alta energía. Casi inmediatamente, el positrón interacciona con un electrón, destruyéndose ambas partículas con la emisión de dos rayos gamma, que se propagan en sentidos casi exactamente opuestos. Las imágenes de TEP se han obtenido partiendo de las mediciones de esta radiación. En ellas, las regiones de alto volumen sanguíneo aparecen coloreadas en rojo, amarillo y blanco y delinean el gran sistema venoso que discurre por la superficie del cerebro. Las zonas coloreadas en azul representan de un modo general el verdadero volumen sanguíneo localizado en el interior de los tejidos cerebrales. Las imágenes de esta ilustración, así como las demás que aparecerán a lo largo de este artículo, han sido obtenidas por los autores en la Facultad de Medicina de la Universidad de Washington.

Tomografía de emisión de positrones

Esta técnica revela la concentración, en el organismo vivo, de sustancias marcadas radiactivamente con mayor definición que la conseguida hasta ahora por cualquier método convencional. Se la considera muy prometedora en biología y medicina

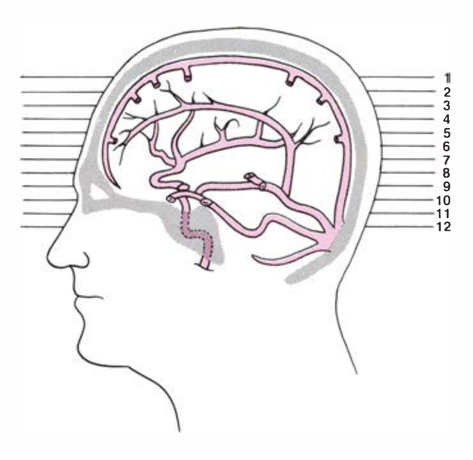
Michel M. Ter-Pogossian, Marcus E. Raichle y Burton E. Sobel

Se admite universalmente, en medicina, que toda actividad biológica es consecuencia de otra bioquímica y que bajo cada tipo de patología subyace un mal funcionamiento de esta última. Es, por tanto, un objetivo permanente de la medicina poder identificar la actividad bioquímica anormal, asociada a una determinada patología, y observarla, con la mayor prontitud, directamente y en el mismo órgano que la sufre. Una técnica muy útil, desarrollada gracias a la posibilidad de obtener isótopos radiactivos artificiales, consiste en administrar al paciente una sustancia marcada radiactivamente y observar el proceso de tránsito de la misma por el interior del organismo mediante un sistema detector adecuado. Los instrumentos convencionales empleados en medicina nuclear para la formación de imágenes proporcionan resultados no exentos de distorsión, siendo el origen de la misma la dependencia existente entre la amplitud y profundidad del campo visual y la inherente incapacidad de discernir la zona de interés de las situadas por delante y por detrás de ella. En esencia, dichos instrumentos se ven obligados, por su principio de funcionamiento, a reducir a dos dimensiones el espacio tridimensional en el que se sitúan los objetos observados. No obstante, las imágenes conseguidas por dichos métodos bastan para muchas aplicaciones clínicas.

En estos últimos años se ha podido localizar con mayor precisión el origen de la radiación emitida gracias al desarrollo de una técnica análoga a la conocida como tomografía axial por computador (TAC), en la cual se construye la imagen de una sección del sujeto haciendo girar una fuente de rayos X alrededor del mismo, mientras un detector situado solidaria y diametralmente opuesto a ella registra las variaciones de absorción sufrida por la radiación al atravesar los tejidos; con los datos obtenidos, un ordenador reconstruye la imagen de cada una de las secciones transversales exploradas. Y resulta evidente que, cuando una situación patológica se pone de manifiesto en una imagen convencional de rayos X, o en un TAC, la situación bioquímica anormal subyacente está, en general, bastante avanzada.

La técnica que hace posible combinar el diagnóstico precoz de un estado bioquímico de carácter patológico, propio de la medicina nuclear, con la localización precisa del lugar en que tal estado sucede, mediante la técnica de reconstrucción de imágenes por ordenador, recibe el nombre de tomografía de emisión de positrones (TEP). En ella, un compuesto químico con la actividad biológica adecuada se convierte en radiactivo mediante la adición de un isótopo emisor de positrones, o electrones positivos, en su proceso de desintegración. El positrón emitido se combina, casi de inmediato, con un electrón, dando por resultado la aniquilación de ambas partículas y la emisión de dos rayos gamma. Estos dos rayos se propagan en sentidos casi exactamente opuestos a través de los tejidos circundantes, siendo registrados, ya en el exterior del sujeto, por un conjunto de detectores distribuidos circularmente alrededor del mismo. La aplicación de un algoritmo matemático adecuado, mediante un ordenador, a la información obtenida por los detectores permite realizar rápidamente la reconstrucción espacial de la intensidad radiactiva, sobre un plano transversal dado del paciente y presentar los resultados sobre una pantalla de rayos catódicos. Los registros se realizan a intervalos de tiempo regulares a partir de la administración del compuesto radiactivo; podemos utilizar en la imagen un código de colores para facilitar la apreciación de diferencias de niveles de actividad de un punto a otro. Convenientemente interpretadas, las imágenes de TEP permiten realizar una estimación, localizada y no invasiva, de numerosos procesos bioquímicos fundamentales para el funcionamiento del órgano explorado.

Hasta el presente, la TEP se ha empleado extensamente con animales de experimentación, perros y monos sobre todo, y en forma más limitada con pacientes humanos; y ha permitido arrojar nueva luz sobre numerosos procesos fisiológicos, imposibles de ser estudiados con parecida eficacia por cualquier otro método. Hace sólo unos pocos años, se consideraba la TEP como un procedimiento exótico y caro, limitado a unos pocos laboratorios; hoy, por contra, está convirtiéndose en una herramienta fundamental para el diagnóstico médico, existente ya en una cuarentena de centros repartidos por todo el mundo.



PLANOS CORRESPONDIENTES a las imágenes de TEP de la página opuesta, definidos sobre una sección del encéfalo. Las estructuras venosas principales, que aparecen destacadas en las imágenes de TEP, están aquí representadas en color.

Además de su capacidad de determinar con precisión la localización de la sustancia marcada y administrada al paciente, la TEP presenta todavía una segunda e importante característica que la sitúa por encima de las técnicas convencionales usadas en medicina nuclear, al ofrecer una selección más útil de elementos químicos, biológicamente significativos, apropiados para servir de marcadores radiactivos. Los sistemas biológicos están constituidos principalmente por compuestos de carbono, nitrógeno, oxígeno e hidrógeno. De los tres primeros existen isótopos radiactivos de vida corta, los cuales se desintegran emitiendo positrones, y son, por ello, apropiados para su aplicación a la tomografía de emisión de positrones. Desde este punto de vista, los isótopos más interesantes son el oxígeno 15, el nitrógeno 13 y el carbono 11, cuyas semividas respectivas son, aproximadamente, de 2, 10 y 20 minutos. A pesar de no existir isótopo alguno de hidrógeno que emita positrones (el tritio, o hidrógeno 3, emite sólo electrones de baja energía), es posible marcar el agua, para aplicaciones de TEP, con oxígeno 15. Con anterioridad al desarrollo de la TEP, el carbono 11, el O-15 y el N-13 veían limitadas sus aplicaciones a investigaciones bioquímicas realizadas con detectores convencionales que registraban las radiaciones γ producidas en el proceso de aniquilación de los positrones emitidos.

El desarrollo de la TEP se ha visto retrasado en gran medida a causa de las cortas vidas medias de los isótopos emisores de positrones de los elementos químicos más interesantes como trazadores. Debido a su efímera existencia, el carbono 11, el oxígeno 15 y el nitrógeno 13 deben producirse muy cerca del punto de detección, y se hace necesario el diseño de técnicas químicas rápidas capaces de introducirlos en los compuestos adecuados. En definitiva, los estudios realizados mediante TEP y compuestos marcados están limitados a procesos bioquímicos rápidos.

En los últimos 20 años, las dificultades asociadas con la producción y procesamiento de isótopos de vida muy corta se han resuelto en su mayor parte. Numerosos centros se hallan hoy equipados con ciclotrones diseñados expresamente para la producción de diversos isótopos emisores de positrones. Dichos ciclotrones aceleran por lo general protones (núcleos de hidrógeno) hasta una energía de 15 millones de electronvolt (MeV) y deuterones (núcleos de hidrógeno pesado) hasta la mitad de dicha energía. Tales instrumentos pesan alrededor de 20 toneladas y están blindados por varias toneladas de hormigón.

La elección de los materiales que sirven de blanco en el ciclotrón depende del isótopo radiactivo deseado y de la forma química en la que aparece. A menudo sirve de tal blanco un gas, con objeto de facilitar el transporte del material activo desde el ciclotrón al lugar donde ha de utilizarse. Incluso cuando el blanco es un sólido, el isótopo obtenido tiene frecuentemente forma gaseosa. En general, el carbono 11 se obtiene bombardeando nitrógeno o boro con protones, el oxígeno 15 por bombardeo de nitrógeno con deuterones y el nitrógeno 13, bien bombardeando oxígeno con protones o carbono con deuterones. Se eligen las reacciones de forma que el isótopo radiactivo resulte relativamente libre de contaminación por isótopos estables del mismo elemento. Tales isótopos "limpios" son altamente deseables en muchos procesos de marcado.

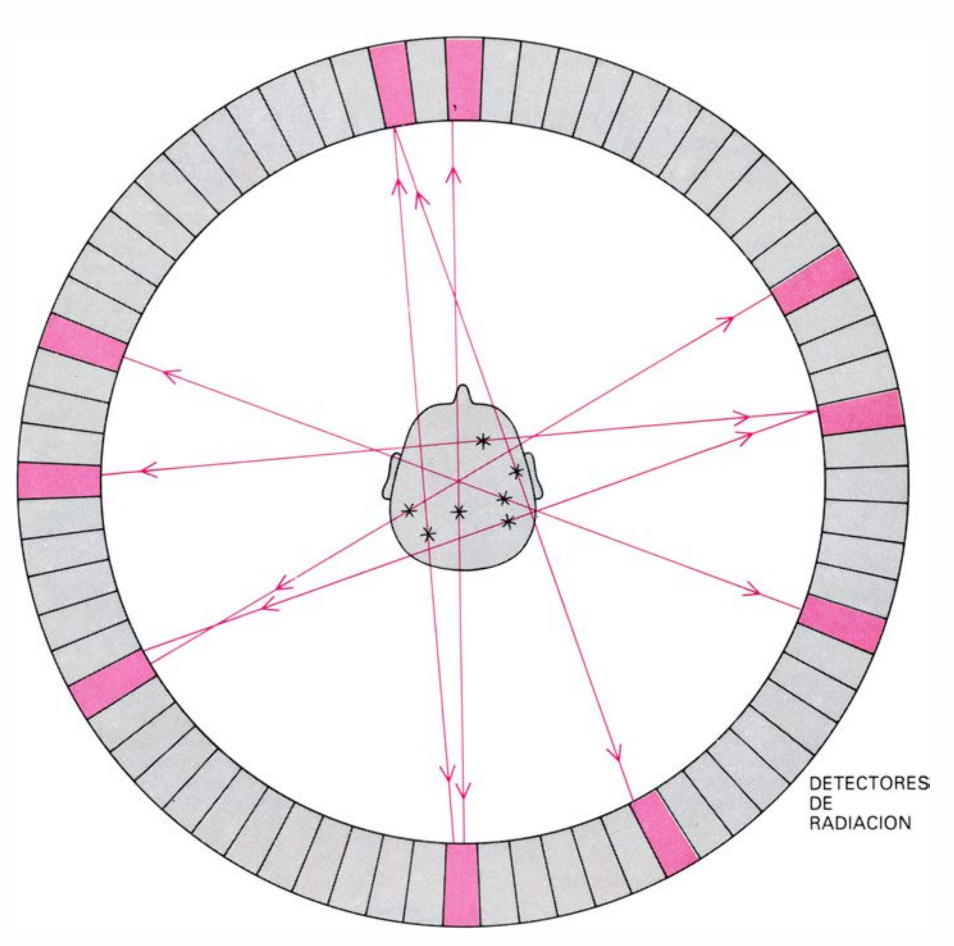
En algunos casos puede usarse directamente el isótopo radiactivo en su forma elemental. Así, el oxígeno 15 sirve para estudios metabólicos y el nitrógeno 13 para estudios de ventilación pulmonar. Varios compuestos sencillos se marcan sin dificultad: agua con oxígeno 15, monóxido de carbono o anhídrido carbónico con carbono 11 o con oxígeno 15 y amoníaco con nitrógeno 13. En otros casos, marcar un compuesto complejo puede detener al químico de mayor imaginación durante meses o años hasta que logra diseñar un proceso de marcado suficientemente rápido para que sea compatible con la vida media del isótopo. Por ejemplo, costó muchísimo marcar glucosa con carbono 11, aun cuando el marcado con carbono 14 (cuya vida media es de 5730 años) estaba ya bien establecido. Se exponen hojas de acelga a la acción simultánea de luz y anhídrido carbónico marcado con carbono 11. A través de la fotosíntesis se incorpora el anhídrido carbónico a la glucosa, que luego se extrae mediante rápidos procesos químicos y se purifica por cromatografía.

Surgen dificultades, a veces, incluso después de haberse desarrollado un proceso rápido de marcado. Por ejemplo, la molécula marcada puede no ser fácilmente soluble en los fluidos del organismo. Tal problema se plantea con la sal del ácido palmítico, un ácido graso utilizado en el estudio de la integridad del músculo cardíaco. Se solubilizó la sal incorporándola a la albúmina del suero proteínico. A veces, el problema está en la obtención de una actividad específica extremadamente alta. Por ejemplo, compuestos adecuados para el estudio de receptores en células nerviosas deben tener una actividad específica de millares de

TOMOGRAFIA AXIAL COMPUTERIZADA, o TAC. Se basa en programas semejantes a los utilizados en la TEP. En una exploración de TAC, al girar la fuente de la radiación alrededor del paciente, se registra la atenuación de los rayos X a lo largo de un gran número de direcciones. En un proceso típico, la fuente de rayos X realiza una exploración lineal transversal en cada una de las 180 posiciones angulares que puede tomar, separadas entre sí un grado. Los datos resultantes de la exploración global permiten que un ordenador reconstruya la opacidad relativa del tejido en una sección transversal plana.

curie por milimol. Esto significa que tiene que marcarse radiactivamente la mayor parte o la totalidad de las moléculas de una muestra. A pesar de las dificultades de varia índole existen en la actualidad métodos para marcar centenares de especies de moléculas con isótopos emisores de positrones, excediendo en número al de las empleadas en medicina nuclear convencional.

Un isótopo particularmente útil en los estudios de TEP es el flúor 18, cuya semivida es de 110 minutos. Además de su obvia capacidad para revelar cuál sea el papel del flúor en la fisiología animal, el flúor 18 tiene una virtualidad adicional: puede incorporarse en moléculas que son análogos de los sustratos naturales del metabolismo. En el interior del organismo, los análogos siguen el proceso metabólico de la molécula normal hasta un cierto punto, donde la secuencia de reacciones se detiene bruscamente porque la maquinaria bioquímica de la célula no admite ya, a partir de ese momento, la molécula del análogo. En el estudio del metabolismo de la glucosa un análogo de ésta, en el que el flúor 18 sustituye a uno de los átomos de oxígeno, se demostró útil.



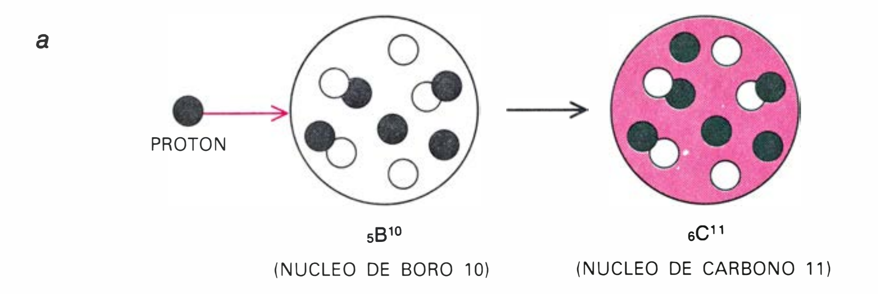
CARACTERISTICA DISTINTIVA DE UNA EXPLORACION DE TEP es que la radiación penetrante tiene su origen en el interior del paciente, en vez de provenir del exterior. Se lleva la fuente de radiación al tejido de interés mediante un compuesto, biológicamente activo, marcado con un isótopo radiactivo que se desintegra emitiendo un positrón. El positrón se aniquila, dentro de un radio de unos pocos milímetros en torno al punto de desintegración, por interacción con un electrón. Los detectores están dispuestos de tal suerte que se registren, exclusivamente, sucesos simultáneos separados entre sí 180 grados. Un programa de ordenador reconstruye posteriormente la distribución de los isótopos en desintegración.

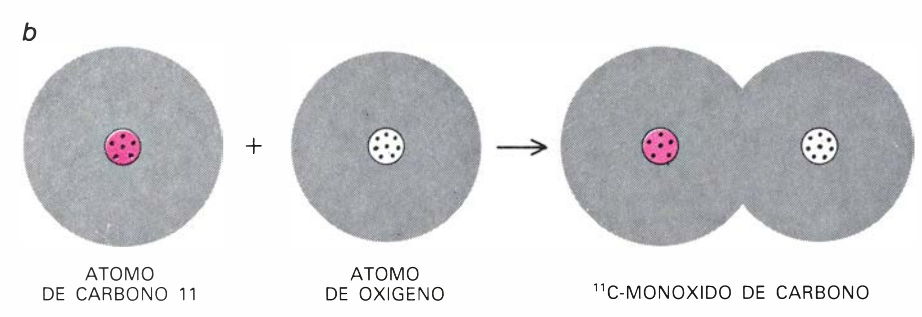
El proceso de formación de la imagen en la tomografía de emisión de positrones consta de tres pasos principales: detección de la radiación gamma emitida en el proceso de aniquilación de positrones, determinación de la dirección de propagación de la radiación y reconstrucción de la distribución de la misma para formar una imagen geométrica exacta. También podría obtenerse una imagen tomográfica con el empleo de isótopos directamente emisores de radiación gamma, pero el número de ellos es limitado. Además, la reconstrucción resulta simplificada y la imagen más fiel si los isótopos emiten positrones. Los positrones emitidos pierden la mayor parte de su energía cinética tras haber recorrido unos pocos milímetros de tejido vivo. Es altamente probable su interacción con un electrón, proceso en que se aniquilan ambas partículas, convirtiéndose sus masas en 1,02 millones de electronvolt de energía, repartidos por igual entre dos rayos gamma, o fotones de alta energía.

Una vez que se han emitido simultáneamente los dos rayos gamma, propagándose en sentidos casi exactamente opuestos, podemos determinar con precisión el lugar de su origen. Ello se consigue asociando en parejas los detectores de radiación, separado cada uno de los elementos de la pareja 180 grados respecto del otro; los dos detectores registran una señal sólo si ambos reciben en coincidencia fotones de alta energía. De esta manera, una distribución circular de detectores es capaz de establecer las fuentes de todos los pares coincidentes de radiaciones gamma que se producen en el interior de un volumen definido por el conjunto de rectas que unen las parejas de detectores; los procesos de aniquilación que tienen lugar fuera de dicho volumen no se registran. De la apertura de los detectores depende la resolución espacial que es posible obtener; queda ligeramente degradada por la penetración de los positrones en la materia y por el hecho de que los fotones emitidos no lo son en sentidos exactamente opuestos.

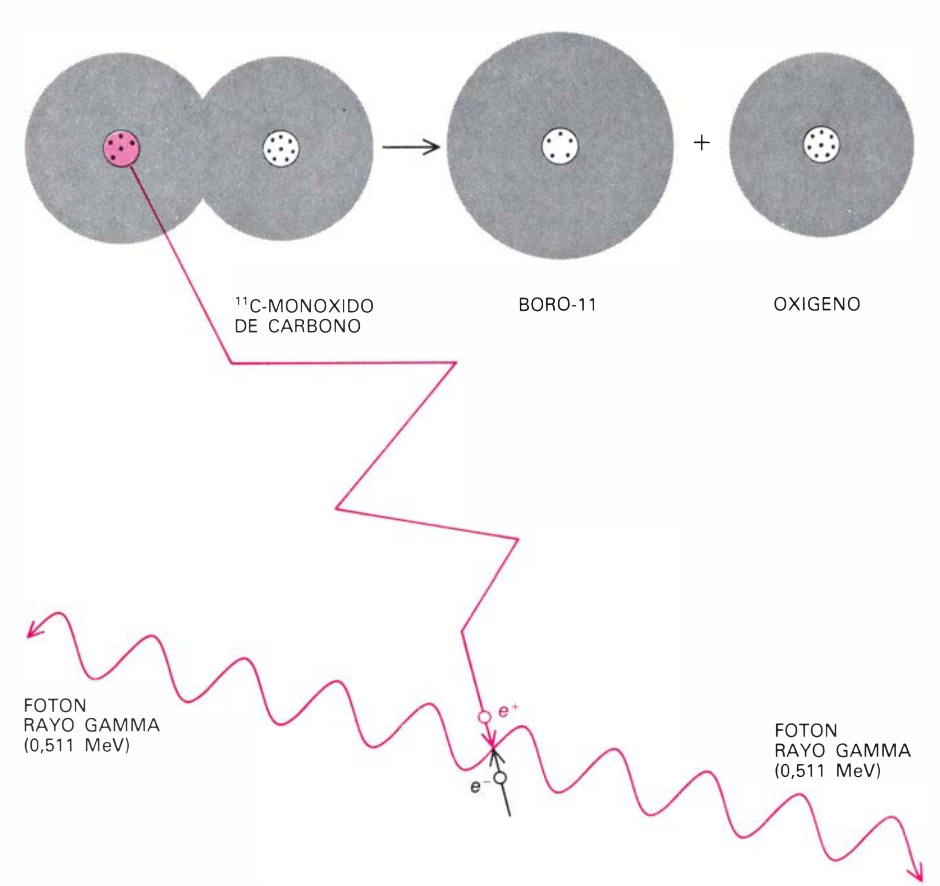
Los sucesos registrados por los detectores de coincidencia constituyen el material bruto a partir del cual se reconstruyen las imágenes de TEP. El principio de la reconstrucción de imágenes por tomografía axial ayudada de ordenador descansa en la posibilidad de reproducir con exactitud un objeto a partir de un conjunto de proyecciones del mismo sobre diferentes ángulos. La fidelidad de tal reconstrucción es proporcional al número de proyecciones; con un número infinito de proyecciones podríamos reconstruir un objeto sin apartarnos un ápice. En un sistema típico de TEP se toman entre 100 y 300 proyecciones, lo que da una resolución especial de algunos milímetros.

Al medir fotones coincidentes, la distribución de detectores en un sistema de TEP permite determinar un conjunto de líneas rectas, situadas todas en un mismo plano. Clasificando este conjunto, se obtiene una distribución de rectas paralelas, muy próximas entre sí, que son representación de la radiactividad del sujeto vista desde ángulos diferentes. Los grupos de rectas paralelas, registradas como perfiles, se someten al proceso matemático llamado convolución; éste dota a dichos perfiles de valores negativos destinados a suprimir las estructuras debidas a los detectores de la imagen. Los perfiles





ISOTOPOS EMISORES DE POSITRONES, originados por bombardeo de elementos no radiactivos con ciertas partículas: protones, deuterones o núcleos de helio. Se obtiene así, por ejemplo, carbono 11 cuando el boro 10, cuyo núcleo consta de 5 neutrones y 5 protones, captura un protón adicional (*a*). En muchos estudios se combina carbono 11 con oxígeno para la formación de ^{11}C-monóxido de carbono (*b*).



DESTINO DE UN COMPUESTO MARCADO en el interior de un organismo al desintegrarse el isótopo radiactivo incorporado para hacer de trazador. Por ejemplo, ^{11}C-monóxido de carbono actúa como trazador de la hemoglobina formando ^{11}C-carboxihemoglobina. Cuando se desintegra un núcleo de carbono 11, libera un positrón y se convierte en boro 11, cuyo núcleo consta de 5 protones y 6 neutrones. Se sigue de ello que el monóxido de carbono queda disociado en un átomo de boro y un átomo de oxígeno. Cuando el positrón (e^+) tropieza con un electrón (e^-), cada uno de los dos rayos gamma, resultantes del proceso de aniquilación, poseen un poco más de medio millón de electronvolt (MeV) de energía.

convolucionados resultantes son "retroproyectados" para dar la imagen final del objeto. Al mismo fin se han aplicado con éxito otros algoritmos no basados en la convolución, pero éste de la "retroproyección filtrada" es la técnica más utilizada en el momento actual.

Una unidad de TEP completa consta de un sistema de adquisición de datos y de un ordenador. El sistema de adquisición consta a su vez de los detectores de radiación, sus circuitos asociados y, en muchos diseños, de un sistema mecánico que comunica un ligero movimiento a los detectores con el fin de realizar un muestreo más eficaz. En algunos diseños, por el contrario, el número de detectores es suficientemente grande para hacer innecesario tal movimiento. La información obtenida por el sistema de adquisición se lleva a un ordenador rápido, que suele poseer incorporada una memoria de gran capacidad. El equipo periférico del ordenador comprende discos y cintas magnéticas, impresora (que proporciona los valores numéricos de la distribución de radiactividad en la imagen), un sistema de presentación para la inmediata visualización y registro de la imagen; con frecuencia posee capacidad interactiva para el análisis de la imagen. Algunos de los sistemas de TEP llevan incorporados los circuitos electrónicos necesarios para realizar partes del proceso de reconstrucción que otros sistemas dejan enteramente al ordenador.

Los sistemas más modernos de TEP pueden registrar simultáneamente hasta 7 imágenes de la cabeza o del tronco humanos con una resolución de alrededor de un centímetro y un error en la distribución de la radiactividad de centésimas, pudiéndose captar imágenes sucesivas a intervalos inferiores a un minuto. La dosis de material radiactivo que se suministra al paciente produce una exposición a la radiación similar a las típicas en otras técnicas de diagnóstico seguidas en medicina nuclear. Hay varias compañías que venden sistemas de TEP a precios que oscilan desde los 600.000 hasta más del millón de dólares.

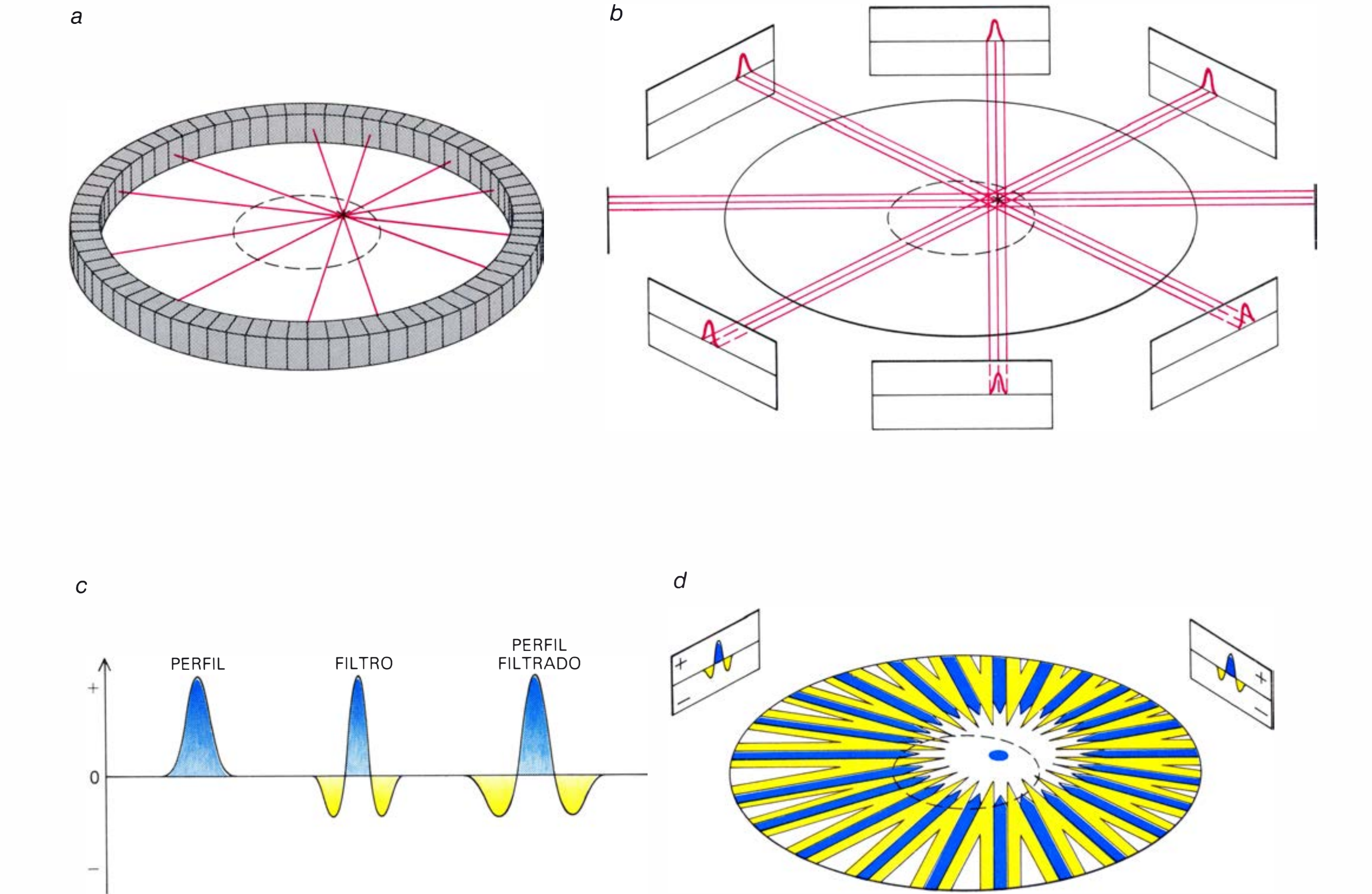
Ocupémonos ahora de las clases de estudios fisiológicos que se pueden llevar a cabo con ayuda de la tomografía de emisión de positrones. Una de las primeras medidas realizadas con su ayuda fue la determinación regional del volumen ocupado por la sangre en cada sección transversal del cerebro humano. Los resultados más satisfactorios se

obtienen haciendo inhalar al paciente aire que contiene trazas de monóxido de carbono marcado con carbono 11. El compuesto ^{11}C-monóxido de carbono se asocia con avidez a la hemoglobina para formar el complejo carbono 11-carboxyhemoglobina, marcando de manera eficaz la totalidad de la sangre del paciente. Se determina fácilmente el volumen sanguíneo cerebral mediante comparación de la cantidad de ^{11}C-carboxyhemoglobina observada en las imágenes tomográficas del cerebro con la existencia en la totalidad del organismo, estimada a partir de muestras de sangre venosa extraídas simultáneamente. Las mediciones del volumen sanguíneo regional del cerebro mediante TEP, además de poner de manifiesto las esperadas diferencias regionales que resultan del hecho de ser mayor la densidad de vasos sanguíneos en la materia gris (que constituye la corteza cerebral) que en la materia blanca subyacente, delinean los vasos sanguíneos principales, especialmente venas, que rodean el cerebro.

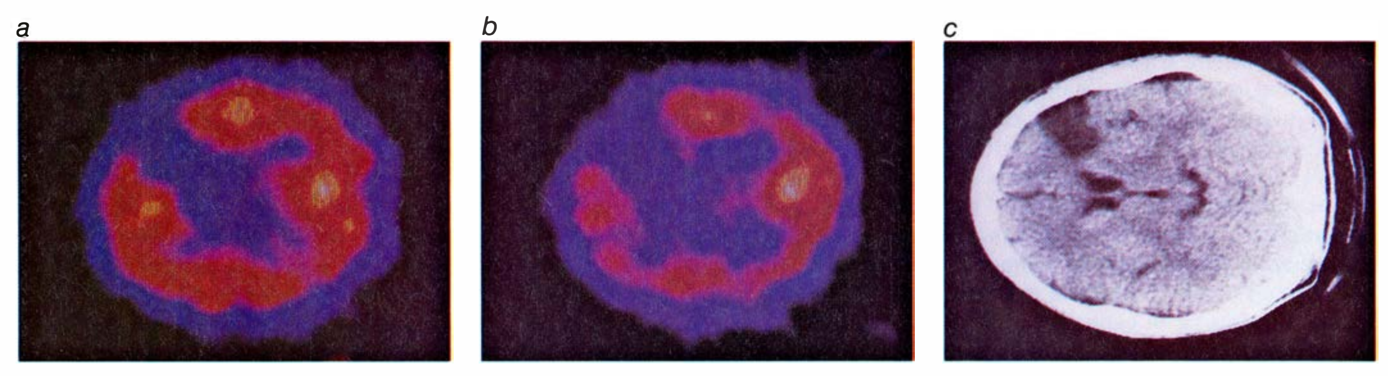
Aunque existen otras numerosas técnicas bien establecidas para medir el volumen sanguíneo cerebral, el éxito del método de TEP revistió gran importancia por varias razones. En primer lugar, demostró que la perfusión regional de un órgano como el cerebro puede determinarse por un método sencillo y exacto. En segundo lugar, la técnica de TEP para medir el volumen cerebral de sangre puede extenderse a otros órganos y tejidos marcando otras sustancias que no sean el monóxido de carbono. Por ejemplo, el marcaje de plaquetas sanguíneas ha permitido, por primera vez, estudiar los procesos de arteriosclerosis en los vasos de un animal vivo. Tal información sería de gran ayuda en el diagnóstico y tratamiento de pacientes humanos que sufren diversas patologías vasculares, incluyendo las oclusiones arteriales, causa de ataques cardíacos.

En tercer lugar, la técnica de medida del volumen sanguíneo por TEP es susceptible de adaptarse a la determinación de la concentración de casi cualquier compuesto adecuadamente marcado, drogas y productos farmacéuticos incluidos, en el interior de los vasos de un órgano considerado, por comparación con su concentración extravascular. La capacidad de realizar tal distinción es exclusiva de la TEP y reviste un enorme interés en la medición del metabolismo y la composición química de tejidos y flujo sanguíneo.

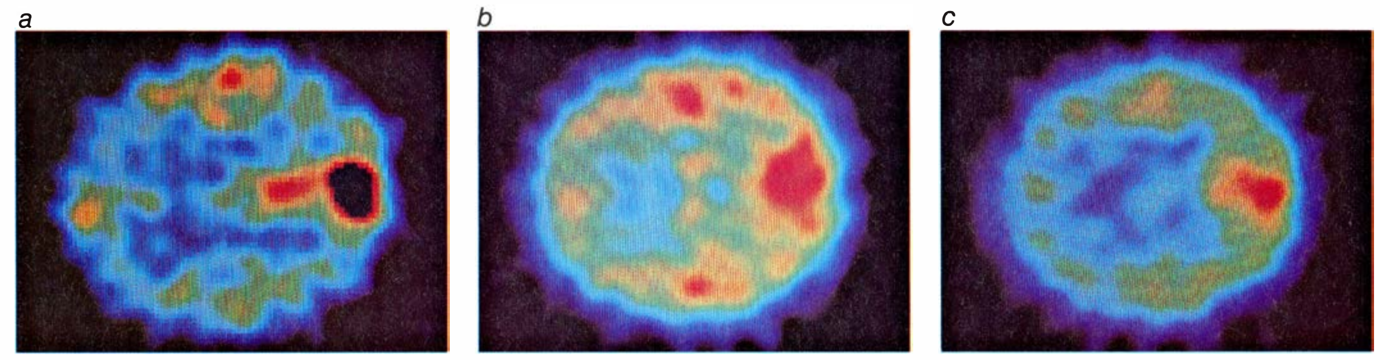
La medición del metabolismo regional es algo más compleja que la medición del volumen sanguíneo regional. Se han realizado estudios cuantitativos de metabolismo cerebral empleando oxígeno marcado con oxígeno 15, glucosa marcada con carbono 11 y el análogo de la glucosa con flúor 18 (^{18}F-2-desoxy-D-glucosa). El metabolismo del músculo cardíaco se ha estu-



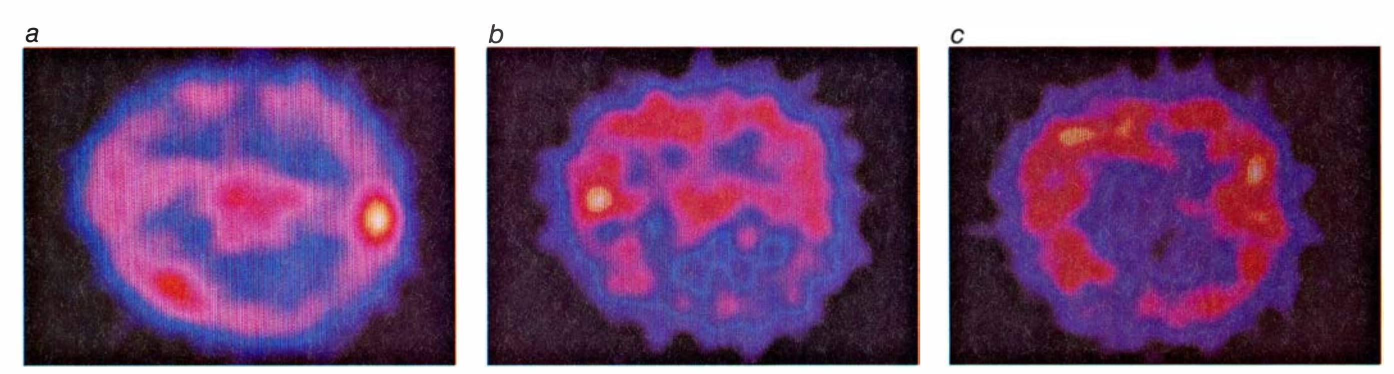
PROCESO DE TEP, que consta de cuatro pasos fundamentales. En el primer paso (*a*) los fotones gamma, que marcan los puntos de desintegración de un compuesto emisor de positrones, se captan por una distribución circular de detectores. Las mediciones representan un conjunto de líneas rectas, situadas todas en un mismo plano. Clasificando convenientemente la información, se obtiene una serie de rectas paralelas (*b*) que representan la proyección, a diferentes ángulos, de la radiactividad existente en el interior de una determinada región del paciente. Estas proyecciones, registradas inicialmente como perfiles, se filtran matemáticamente, mediante un proceso llamado de convolución, que evita la aparición de artefactos en la imagen (*c*). Los perfiles, después de convolucionados, se "retroproyectan" para reconstruir la imagen de la distribución de la sustancia marcada (*d*). El procesamiento de la gran cantidad de datos recogidos exige un ordenador rápido. La TEP se practica ya en una cuarentena de centros repartidos por todo el mundo.



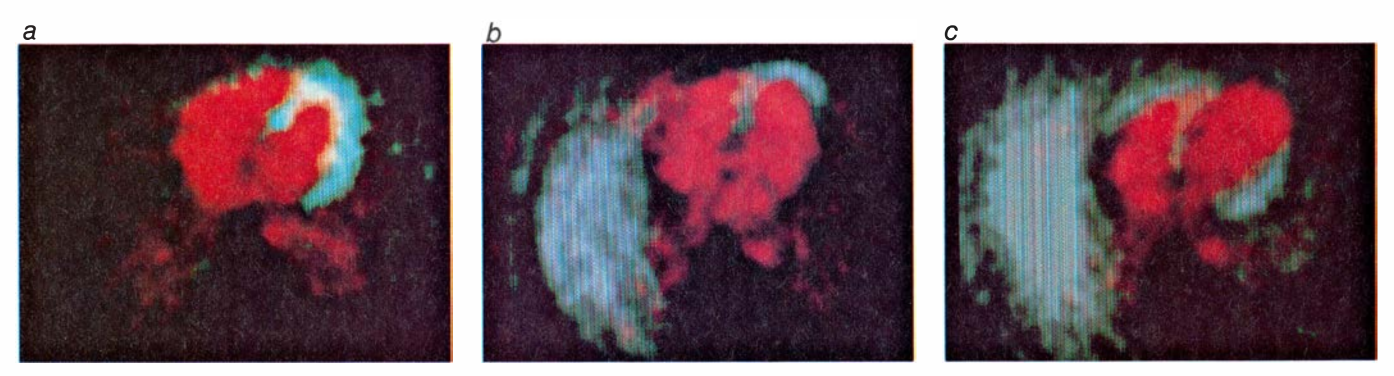
IMAGENES DE UN INFARTO CEREBRAL obtenidas mediante exploraciones de TEP (*a* y *b*) y de TAC (*c*). La región dañada está situada en el cuadrante superior izquierdo (con el paciente mirando hacia la izquierda). Tales infartos aparecen inmediatamente en una exploración de TEP. La exploración de TAC revela sólo variaciones de la densidad del tejido. La primera imagen TEP (*a*), realizada después de que el paciente recibiera una muestra de sangre marcada con ^{15}O-agua, corresponde al flujo sanguíneo cerebral. La segunda de las imágenes (*b*) muestra el consumo de oxígeno.



LOCALIZACION DE UN FOCO EPILEPTICO en imágenes del TEP volumen de sangre cerebral (*a*), flujo de sangre (*b*) y consumo de oxígeno (*c*). En la imagen *a* la pequeña marca roja, situada en la posición de las 12 del reloj, representa un incremento del volumen sanguíneo en el lugar del ataque.



LOCALIZACION DE UN TUMOR CEREBRAL, un meningioma de grandes dimensiones, identificable en estas imágenes de TEP, que corresponden también al volumen de sangre cerebral (*a*), flujo de sangre (*b*) y consumo de oxígeno (*c*). Los meningiomas son tumores altamente vascularizados, situados en las membranas que envuelven el cerebro. En este caso su presencia pone de manifiesto por un incremento inusitado del volumen de sangre en la región central. Se aprecia también un aumento local del flujo sanguíneo así como una disminución del consumo de oxígeno en el tumor.

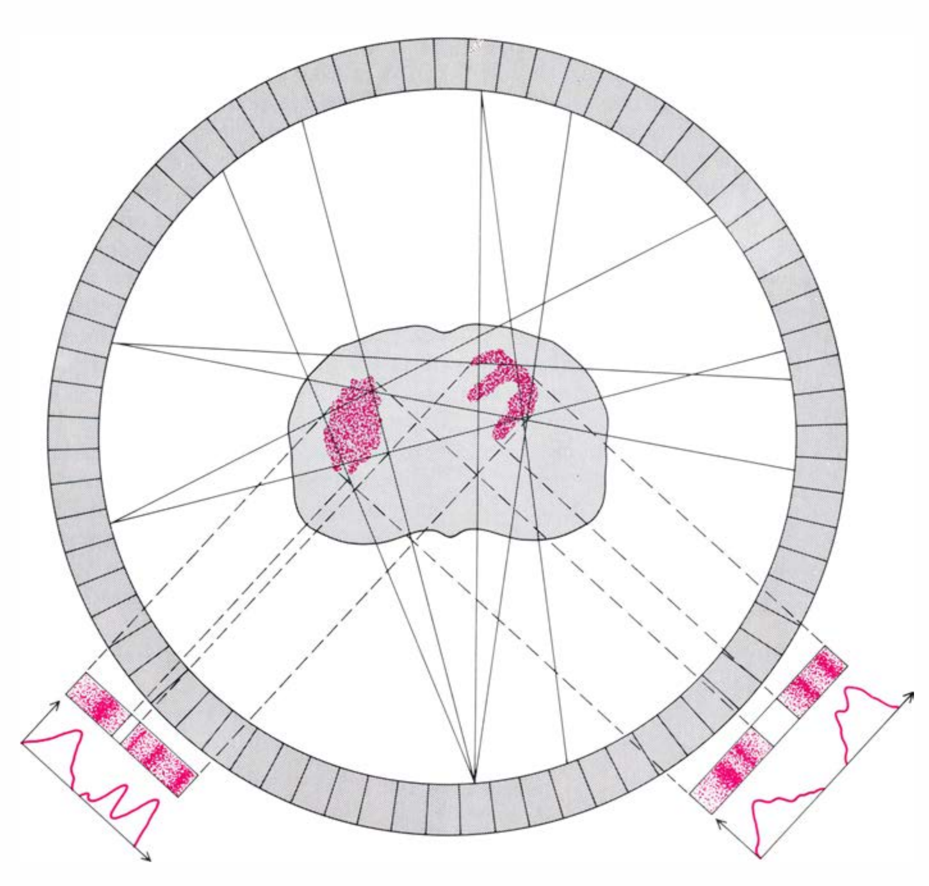


UN CORAZON NORMAL Y UN CORAZON DAÑADO presentan diferencias notables en estas imágenes de TEP, que corresponden a la respuesta del corazón a dos sustancias trazadoras. Las zonas sombreadas en verde muestran la distribución del isótopo radiactivo carbono 11 inyectado en forma de ^{11}C-palmitato, sal del ácido palmítico, uno de los ácidos grasos de metabolización rápida. Las zonas sombreadas en rojo corresponden a la distribución de carbono 11 inhalado en forma de ^{11}C-monóxido de carbono. En un corazón normal (*a*), los ventrículos se distinguen por su color verde brillante. Las zonas verdes aparecen truncadas en la imagen correspondiente a un paciente con infarto lateral de miocardio (*b*) y a un paciente con infarto anterior de miocardio (*c*). La estructura de color verde de gran tamaño que podemos apreciar a la izquierda del corazón corresponde al hígado.

diado con palmitato, la sal del ácido palmítico y sustrato fisiólogico primario del corazón, marcado con el isótopo carbono 11.

Recientemente, se han desarrollado numerosas técnicas prometedoras para el diagnóstico cardiológico no invasivo. Entre ellas se cuentan la determinación de las dimensiones de las cámaras cardíacas y del movimiento de las válvulas mediante ultrasonidos (ecocardiografía), la evaluación de la función de bombeo del corazón mediante isótopos radiactivos administrados por vía intravenosa (radioventriculografía) y la definición de la anatomía cardíaca por TAC. Estos métodos y otros posibles, sin embargo, se manifiestan intrínsecamente incapaces de determinar, cuantitativa y definitivamente, la perfusión y el metabolismo regionales del miocardio. El problema cardíaco más frecuente entre adultos es la isquemia, trastorno producido por un insuficiente riego sanguíneo de ciertas regiones del corazón. Los síntomas se extienden desde los dolores pectorales episódicos (angina pectoris) hasta la insuficiencia circulatoria (fallo cardíaco congestivo) y el infarto de miocardio (ataque cardíaco).

La enfermedad de la arteria coronaria, responsable de estos trastornos, da lugar no solamente a síntomas sino a desarreglos en el metabolismo regional interno del corazón. Con la técnica de TEP, trazadores tales como el carbono 11 pueden localizarse con precisión y exactitud en el interior del músculo cardíaco. Además, cuando se añaden trazadores emisores de positrones a sustancias adecuadas, tales como el palmitato, se posibilita la caracterización del metabolismo regional con una sustancia cuyo comportamiento químico es idéntico al de la sustancia fisiológica cuyas huellas se están siguiendo.



RETROPROYECCION DE LOS PERFILES CONVOLUCIONADOS a fin de reconstruir la distribución de radiactividad en el interior del paciente en una sección determinada del cuerpo. Aquí la sección elegida pasa a través del corazón y del hígado; la columna vertebral del paciente está en la parte inferior. De acuerdo con las convenciones admitidas en TEP, el punto de vista está situado a los pies, de modo que la estructura en forma de *W* situada a la derecha es el corazón, con el hígado a la izquierda. La distribución resultante es típica de las obtenidas cuando se suministra al paciente una inyección de ^{11}C-palmitato. El palmitato, sal del ácido palmítico, se transporta al corazón y al hígado, a los que suministra energía.

Al objeto de definir la naturaleza externamente detectable de los principales procesos bioquímicos que tienen lugar en el corazón y son susceptibles de determinar mediante TEP, se han realizado estudios en corazones aislados a partir de animales de experimentación y en corazones de pacientes humanos vivos. La acumulación de un cierto trazador en una región del corazón depende de numerosos factores, entre los que se cuenta la velocidad con que se suministra el trazador al tejido (función de la perfusión regional), el tiempo de permanencia del trazador, la fracción del mismo que se extrae de una vez a través de la circulación coronaria y las velocidades de metabolismo y desaparición del trazador inicialmente extraído por el músculo cardíaco. La fracción extraída y el metabolismo dependen críticamente no sólo de la naturaleza de la sustancia rastreada, sino también del estado del metabolismo regional del músculo cardíaco en el momento de su evaluación.

Se llegó a esa crítica relación a través de ensayos con corazones aislados a partir de animales y perfundidos luego con suero nutritivo. Segundos después de inyectar en el suero palmitato marcado con carbono 11, la imagen de TEP registra un pico de actividad proporcional a la cantidad de trazador inyectado. El pico de actividad se atenúa rápidamente al ser arrastrada una fracción importante del trazador por la circulación coronaria. A esta atenuación rápida sigue otra más lenta, debida principalmente al intercambio y eliminación del trazador del suero entre los vasos sanguíneos. La fase tercera y final del proceso de atenuación de la actividad presenta una lenta variación exponencial y refleja la incorporación del carbono 11-palmitato a los lípidos de las células del miocardio.

Comparando esta fase final con la subida inicial de la curva de radiactividad en función del tiempo, se obtiene un índice del tanto por ciento neto de incorporación del ^{11}C-palmitato a los lípidos neutros. En estudios realizados con corazones perfundidos de animales se controlan separadamente las exigencias de flujo y metabolismo del corazón. En condiciones de flujo constante y metabolismo variable, la velocidad de atenuación de la radiactividad del corazón al final de la curva de radiactividad en función del tiempo refleja cambios cuantitativos en el metabolismo del músculo cardíaco. Tales observaciones sugirieron que una disminución en el suministro de ácido graso libre producida por una isquemia debía detectarse fácilmente mediante TEP, tanto en animales de experimentación como en pacientes humanos.

Cuando se administra ^{11}C-palmitato por vía intravenosa a animales de experimentación, se acumula homogéneamente en el miocardio. Cuando éste sufre una isquemia, en cambio, las imágenes de TEP presentan una marcada disminución en la absorción de

^{11}C-palmitato. La distribución del palmitato y la extensión de los defectos de absorción en las imágenes se corresponden estrechamente con las observadas en exámenes post-mortem, determinadas por criterios histológicos y enzimáticos de daño irreversible del miocardio.

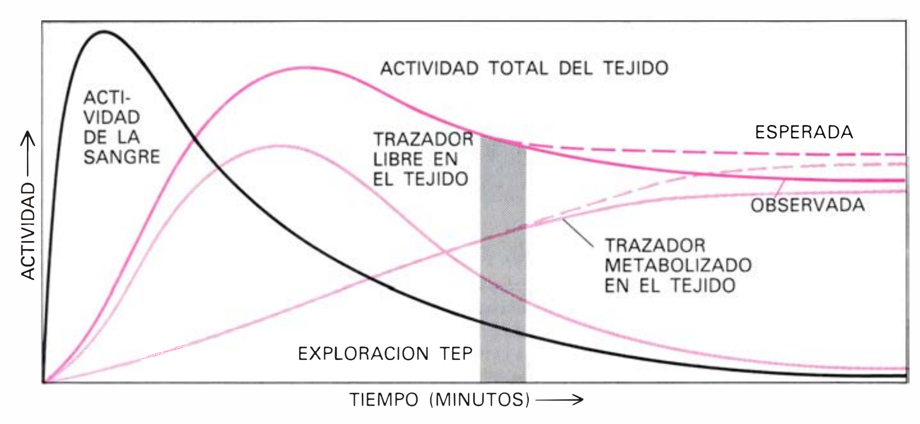
Además de poner de manifiesto los daños irreversibles del corazón, los estudios realizados mediante TEP y ^{11}C-palmitato permiten delimitar zonas deprimidas del miocardio, donde el riego de las arterias coronarias ha sufrido una oclusión transitoria. Las investigaciones tomográficas secuenciales permiten establecer diferencias entre daños reversibles e irreversibles: el miocardio dañado de forma reversible es incapaz, inicialmente, de acumular ^{11}C-palmitato, pero lo hace de forma normal cuando se ha restablecido el flujo sanguíneo.

En individuos que no padecen trastornos declarados de la arteria coronaria ni síntomas de ello, las imágenes de TEP demuestran que la distribución de ^{11}C-palmitato, administrado por vía intravenosa, es homogénea en todo el ventrículo. En cambio, en pacientes que han sufrido de infarto de miocardio en el pasado (lo que se traduce en la muerte de un cierto número de células del miocardio), las imágenes de TEP revelan zonas de absorción reducida de ^{11}C-palmitato, que coinciden con las que sufrieron el infarto (determinadas por medio de electrocardiografía). Además, el alcance del infarto sufrido por estos pacientes y detectado por TEP se corresponde exactamente con la gravedad determinada por procedimientos bioquímicos.

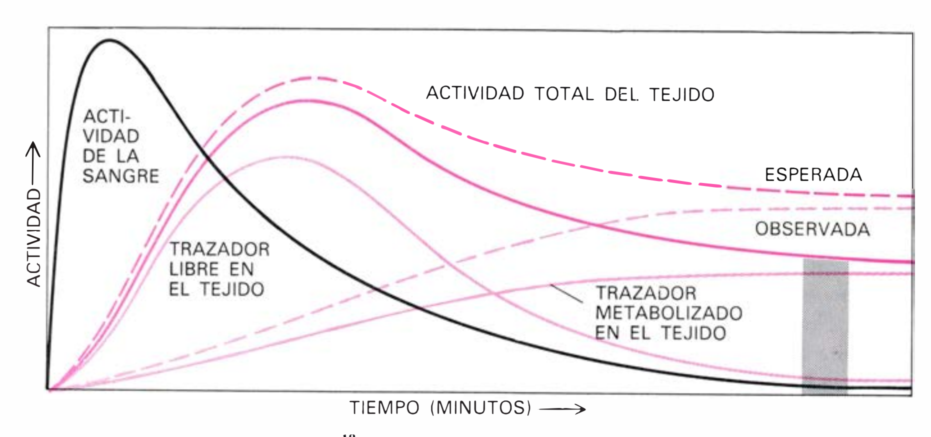
La tomografía de emisión de positrones con ^{11}C-palmitato ha colaborado, también, en el esclarecimiento del mecanismo responsable del trastorno global, o generalizado, de la función de bombeo del corazón en pacientes que sufren de infarto agudo de miocardio. De acuerdo con cierto punto de vista, el comportamiento anormal de algunas regiones de la pared cardíaca se deben principalmente a un retardo en la conducción de los impulsos eléctricos que excitan a dicho músculo. Otra de las hipótesis existentes supone que dichas anormalidades son reflejo de disfunciones bioquímicas y mecánicas de la pared cardíaca. En los pacientes estudiados por nosotros, el deterioro global de la función ventricular y la localización de anormalidades en el movimiento de la pared cardíaca están, ambos, estrechamente relacionados con un metabolismo trastornado del miocardio, puesto de manifiesto por TEP. En estos pacientes resulta evidente que al menos las anomalías de funcionamiento son atribuibles, en primer lugar, al trastorno metabólico regional del miocardio, inducido más por una perfusión disminuida, que por anormalidades en la conducción.

Los ensayos con ^{11}C-palmitato proporcionan un ejemplo de las posibilidades que posee la TEP para determinar el metabolismo regional del músculo cardíaco. Otros agentes, tales como la desoxiglucosa marcada con flúor 18, se han utilizado para la obtención de imágenes cualitativamente semejantes a las obtenidas mediante ^{11}C-palmitato. Existe, no obstante, cierta dificultad en la interpretación exacta de las imágenes obtenidas con flúor 18, ya que los compuestos fluorados presentan con frecuencia propiedades metabólicas distintas de las correspondientes a los metabolitos naturales.

El hecho de que los trazadores utilizados en la TEP presenten un comportamiento complejo en el organismo suele obligar a recurrir a modelos matemáticos y procesos de adquisición de datos muy elaborados, que permiten obtener los valores de las velocidades metabólicas a partir de los datos numéricos suministrados por las imágenes de TEP, juntamente con los procedentes de las medidas de radiactividad realizadas en muestras de sangre arterial o venosa, extraídas del brazo o la pierna del paciente. Numerosas e importantes exigencias se imponen a



INTERPRETACION DE LAS MEDIDAS REALIZADAS POR TEP. Dicha interpretación depende del uso de modelos matemáticos que relacionan la concentración de compuestos trazadores determinados, observados mediante TEP en lugares específicos dentro de un tejido, con un proceso metabólico determinado. Se necesitan dos mediciones simultáneas: el nivel de actividad del trazador en la sangre, que se calcula a partir de muestras de una arteria, o una vena, periférica y la actividad en el tejido explorado, determinada mediante TEP. La actividad del tejido consta de dos partes, correspondientes al trazador metabolizado y al no metabolizado. Dado que la técnica TEP no permite distinguir entre ambas, se hace crítico el modelo matemático a utilizar. Por ejemplo, cuando se mide el metabolismo con un substrato marcado, ^{11}C-glucosa por ejemplo, importa medir en seguida la actividad del tejido para evitar un error respecto al valor esperado, causado por la pérdida de trazador metabolizado por parte del tejido. Gran parte de la actividad medida pertenece al trazador no metabolizado, incluso cuando se ha tenido en cuenta la actividad de la sangre en la región de interés. Un buen modelo matemático debe permitir al usuario de la TEP distinguir entre el trazador metabolizado y el libre y establecer el metabolismo local.

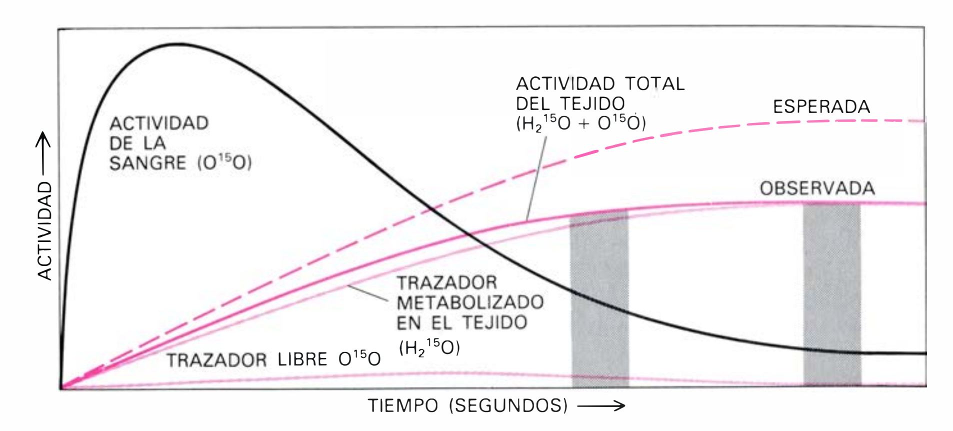


METABOLISMO DEL ANALOGO ^{18}F-2-desoxi-D-glucosa, que difiere en dos importantes aspectos del de la glucosa normal: no se absorbe tan ávidamente por las células del tejido, y, una vez absorbido, no se libera en absoluto. En razón de esta última propiedad podemos retrasar las mediciones de TEP, reduciendo de esta manera la contribución radiactiva del trazador que se halla libre en la sangre y en el tejido. Aún así, las células no aceptan al análogo con la misma facilidad que si se tratara del metabolismo normal, y en consecuencia, la actividad que se observa en el tejido difiere del valor esperado.

tales medidas. Se supone, en primer lugar, que la sustancia marcada se transporta y metaboliza del mismo modo y a la misma velocidad que la sustancia no marcada que se investiga. Esta exigencia se cumple cuando se marca con carbono 11 una sustancia como la glucosa. No se cumple, sin embargo, con sustancias análogas en las cuales, por ejemplo, el flúor 18 sustituye al oxígeno, siendo preciso aplicar factores de corrección adecuados.

Una segunda exigencia consiste en la necesidad de que el trazador metabolizado permanezca en el interior del órgano investigado durante el tiempo de medición. Con un sustrato como glucosa marcada con carbono 11 podemos esperar su permanencia en el cerebro, por ejemplo, durante el tiempo necesario para realizar una exploración. Inmediatamente después de penetrar en las células cerebrales, la ^{11}C-glucosa es fosforilada, reacción que produce ^{11}C-glucosa-6-fosfato. Por ser el cerebro pobre en enzimas capaces de invertir esta reacción e impermeable a la glucosa-6-fosfato, el carbono 11 queda retenido hasta que la ^{11}C-glucosa-6-fosfato se transforma en una serie de subproductos, que acaban por abandonar el cerebro arrastrados por el flujo de sangre venosa. En virtud de ese proceso, el carbono 11 permanece retenido en el órgano durante tiempo suficiente (generalmente 5 minutos) para que pueda medirse el metabolismo a través de las imágenes de TEP. Cuando se emplean análogos tales como glucosa marcada con flúor 18, los requerimientos de tiempo son mucho menores, ya que el metabolismo se detiene con la fosforilación. El fosfato marcado con flúor resta retenido de forma irreversible en las células del sistema nervioso.

Una tercera exigencia en estudios metabólicos realizados con técnica de TEP consiste en que la cantidad de trazador no metabolizado, y por tanto libre en la sangre y en el fluido extracelular, debe ser despreciable o exactamente conocido durante el tiempo de medición. Esta condición se cumple sin mayores dificultades cuando los trazadores son análogos radiactivos. Dado que tales trazadores quedan irreversiblemente retenidos en el órgano estudiado (y en células situadas en cualquier otro lugar), se nos permite retrasar la exploración hasta que los trazadores libres en la sangre y tejidos se atenúen hasta niveles insignificantes. Cuando los trazadores son análogos de glucosa, basta con un tiempo de espera entre 30 minutos y 1 hora. Si los trazadores no son análogos, sino verdaderos sustratos metabólicos, tal espera no es posible, ya que lo metabolitos marcados comienzan a abandonar el tejido unos cinco minutos después del momento de su inyección. Al decidir la forma práctica de realizar la exploración, se debe, por tanto, tener explícitamente en cuenta la existencia de trazador libre en sangre y tejido. Esto se consigue, en parte, midiendo el volumen regional de sangre junto con el metabolismo regional. Conocido el volumen regional de sangre, podemos calcular entonces la cantidad de trazador libre, no sólo en ésta, sino también en el propio tejido.



MEDICION DEL CONSUMO DE OXIGENO, cuya interpretación requiere de otro modelo matemático. Los tejidos con un metabolismo activo, el cerebro por ejemplo, no almacenan oxígeno casi en absoluto; por eso, el oxígeno radiactivo (^{15}O) se convierte rápidamente en agua ($H_2{}^{15}O$). De ahí que la medición de TEP, corregida de la actividad sanguínea, corresponda sólo al trazador metabolizado. Como el agua abandona rápidamente el tejido, la actividad observada decrece de un modo continuo por debajo del valor esperado. Una forma de superar este escollo y aumentar la exactitud de la medida del consumo de oxígeno consiste en realizar numerosas exploraciones secuenciales, a intervalos de pocos segundos.

Cabe optar por un método alternativo al procedimiento anterior de calcular la cantidad de trazador libre en sangre y tejido; se trata de seguir la salida del trazador metabolizado del órgano estudiado, por repetición de la medición del metabolismo varias veces durante los primeros cinco minutos inmediatos a la inyección de dicho trazador. La velocidad metabólica calculada a partir de estas mediciones sucesivas debe, obviamente, decrecer con el tiempo; a partir de la velocidad de decrecimiento, podemos saber cuál sea la verdadera velocidad metabólica. En estudios anteriores nos pareció tal procedimiento bastante realizable. Comprobamos este mecanismo tanto con ^{11}C-glucosa como con ^{15}O-oxígeno. Esta última prueba resulta particularmente exigente debido a la velocidad a la que el oxígeno 15 se metaboliza en agua y desaparece del tejido. En ambos casos, obtuvimos excelentes resultados de la velocidad real a la que se utilizan las sustancias marcadas. En los primeros experimentos nos limitamos a imágenes de un único plano del tejido. Hoy se obtienen resultados mejores todavía con sistemas de TEP diseñados para explorar simultáneamente varios planos.

Con estos rápidos sistemas TEP de planos múltiples resulta también más fácil determinar la cantidad de trazador no metabolizado que permanece en el tejido. Los datos necesarios para ello pueden recabarse de una secuencia de imágenes correspondientes a la actividad regional del tejido en función del tiempo, durante el transcurso de una inyección lenta de trazador por vía intravenosa. Si este procedimiento resulta, como se espera, conseguirá saberse algo más que la mera velocidad de utilización local de un sustrato, como glucosa o palmitato. Se podrá determinar la concentración de sustrato en el órgano investigado, su distribución espacial y la duración de su permanencia en él. Tratándose de pacientes humanos, tales determinaciones serán de extraordinario valor para el conocimiento definitivo de procesos patológicos, así como su diagnóstico y tratamiento.

A pesar de que los estudios de metabolismo realizados con análogos marcados radiactivamente, en lugar de verdaderos sustratos, tienen la ventaja de que quedan retenidos de forma irreversible después de metabolizados, no dejan de presentar sus inconvenientes. Los análogos no son totalmente idénticos, desde un punto de vista bioquímico, a los sustratos naturales; y es preciso determinar experimentalmente las diferencias respecto a la velocidad normal del metabolismo.

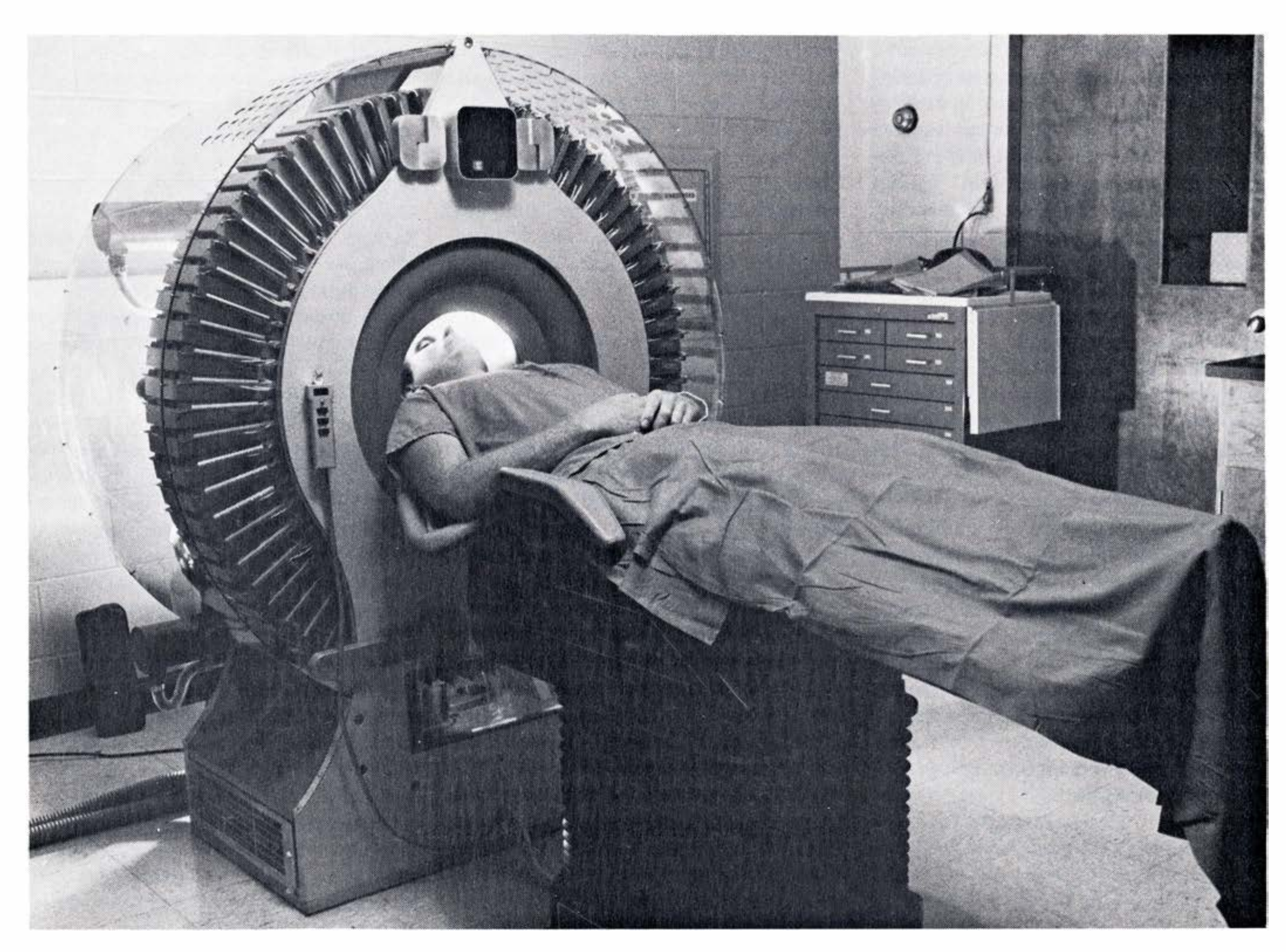
Existen, además, importantes diferencias en el metabolismo de los análogos según la especie animal de que se trate. Y, por si fuera poco, son de esperar nuevas diferencias cuando se trata de estudiar un órgano enfermo. Otras dificultades surgen en relación con el tiempo y con las dosis de radiación. Tratándose de análogos de sustratos, es preciso dejar transcurrir un período de 30 minutos o una hora antes de realizar las mediciones mediante TEP, al objeto de eliminar el trazador no metabolizado de la sangre y el tejido.

Estas exigencias de tiempo dificultan el estudio de ciertos fenómenos transitorios, los ataques epilépticos por ejemplo, cuya duración puede ser de sólo unos pocos minutos. Ello impone también el uso de trazadores de larga vida; y, si éste es de corta vida, obliga a administrar al paciente una dosis de radiación bastante elevada. Finalmente, los análogos de sustratos pueden provocar trastornos en el metabolismo celular normal si se administran en cantidades excesivas. Se debe procurar, por tanto, que el análogo marcado radiactivamente esté relativamente libre de material no marcado.

Además de la determinación de las velocidades de procesos continuos como puede ser el metabolismo, la técnica de TEP permite medir cuantitativamente la composición química de los tejidos si se eligen adecuadamente los compuestos marcados, que se distribuyen rápidamente entre la sangre y el órgano de interés. Así, se ha medido el contenido de anhídrido carbónico del tejido cerebral del mono rhesus, haciéndole inhalar anhídrido carbónico marcado con carbono 11. Las medidas se realizaron utilizando distintas concentraciones de anhídrido carbónico en la sangre y pusieron de manifiesto la esperada relación entre los niveles de dicho gas en la sangre arterial y en el cerebro. Los experimentos sirven también para ilustrar las posibilidades únicas que posee la TEP en lo concerniente a la determinación cuantitativa de acidez del tejido cerebral vivo. Esas mediciones de acidez resultarán de valor a la hora de apreciar los daños producidos por ataques cerebrales.

El proceso general de medida del anhídrido carbónico en el cerebro puede extenderse a la exploración de la respuesta de casi cualquier órgano a cualquier producto farmacéutico que pueda marcarse con un isótopo emisor de positrones. Una aplicación inmediata es la medición de la concentración de productos químicos en órganos. Estudios realizados con TEP ayudarán a definir, por ejemplo, la eficacia terapéutica de los anticonvulsivos en el cerebro y de la digitalina en el corazón. Se podrá determinar en pacientes humanos, por primera vez, la respuesta de receptores y transmisores cerebrales específicos a productos químicos con características de enlace específicas. Tales investigaciones se han practicado en animales. Existen razones para esperar que investigaciones similares, realizadas en humanos, sirvan para poner en claro la acción de agentes psicotrópicos, en condiciones para las cuales no existen modelos satisfactorios obtenibles de animales, como la esquizofrenia y la enfermedad de Parkinson.

El desarrollo de un método de medi-



ULTIMO MODELO de sistema de TEP, el PET VI, diseñado y construido en la Facultad de Medicina de la Universidad de Washington. La estructura cilíndrica alberga los detectores de protones, que registran el paso de la radiación gamma, procedente de la aniquilación de pares positrón-electrón.

ción, completamente satisfactorio, del flujo sanguíneo mediante TEP se ha retrasado con relación a otras aplicaciones. Y ello, por múltiples causas. En primer lugar la fracción de extracción de muchos trazadores está influida por el estado metabólico del tejido y por el tiempo de permanencia, lo que dificulta la cuantificación exacta de la perfusión regional. En segundo lugar, la retención incompleta del trazador inicialmente extraído por el tejido investigado y la recirculación del mismo oscurecen la relación entre acumulación y perfusión regional. En tercer lugar, la absorción del trazador por elementos liposolubles del tejido puede diferir de la correspondiente a los elementos hidrosolubles, lo que contribuye a hacer variable la relación entre perfusión regional y trazador acumulado en el interior de una región determinada del órgano.

Sumándose a estas complejidades biológicas, la determinación cuantitativa de la perfusión regional de órganos tales como el corazón y el cerebro con isótopos radiactivos se ha visto limitada por dificultades encontradas en la determinación exacta de la cantidad de trazador existente en regiones específicas del órgano. La TEP promete la superación de tales dificultades; y así, se ha conseguido la perfusión regional del corazón con la inyección intravenosa de isótopos emisores de positrones seleccionados y la detección de su distribución mediante tomografía. Un trazador empleado para este fin es el ^{13}N-amoníaco ($^{13}NH_3$), compuesto cuyo comportamiento se cree afín al del potasio y otros cationes monovalentes. Otro trazador en el que se tienen puestas las esperanzas es el rubidio 81, más semejante al potasio. El amoníaco se metaboliza activamente y se incorpora a la glutamina por el miocardio, de modo que su fracción de extracción no es constante, sino que depende de factores metabólicos.

En corazones de conejo aislados y perfundidos, la retención del trazador ^{13}N-amoníaco resulta notoriamente influida por el estado metabólico del corazón en condiciones en las que el flujo puede controlarse con rigurosa precisión. A pesar de tales limitaciones, se ha propuesto que la reducida acumulación de ^{13}N-amoníaco en el músculo cardíaco proporciona una estimulación cuantitativa de una perfusión regional disminuida. La acumulación de trazador puede resultar, no obstante, de diversos fenómenos contradictorios: decrecimiento en la fracción de extracción que acompaña a una isquemia regional y resultante de un metabolismo alterado del miocardio, posibles aumentos en la fracción de extracción que reflejan un tiempo de permanencia prolongado en regiones de bajo flujo sanguíneo, posibles cambios de permeabilidad en zonas isquémicas que influyen sobre la fracción de extracción y una liberación disminuida de trazador que evidencia ella misma el bajo flujo sanguíneo. La complicada interacción de estos y otros fenómenos relacionados con ellos puede limitar la utilidad del ^{13}N-amoníaco para definir cuantitativamente la perfusión regional, aun cuando pueda detectarse su distribución mediante TEP.

MARCADOR	SEMIVIDA	SUSTANCIA MARCADA	UTILIZACION
OXIGENO 15	~ 2 MINUTOS	OXIGENO	METABOLISMO
		AGUA	FLUJO SANGUINEO
		MONOXIDO DE CARBONO	VOLUMEN SANGUINEO
		ANHIDRIDO CARBONICO	FLUJO SANGUINEO
NITROGENO 13	~ 10 MINUTOS	AMONIACO	PERFUSION EN ORGANOS
			METABOLISMO
		VARIOS AMINOACIDOS	METABOLISMO
			PERMEABILIDAD (CEREBRO)
		OXIDO NITROSO	FLUJO SANGUINEO
		1,3 BIS (2-CLOROETILO) 1-NITROSOUREA (BCNU)	NIVEL DE DROGAS EN TUMORES
CARBONO 11	~ 20 MINUTOS	MONOXIDO DE CARBONO	VOLUMEN SANGUINEO
		ANHIDRIDO CARBONICO	*p*H DE LOS TEJIDOS
		VARIOS ALCOHOLES	PERMEABILIDAD (CEREBRO)
			FLUJO SANGUINEO
		VARIOS ETERES	FLUJO SANGUINEO
			CONTENIDO DE LIPIDOS EN TEJIDOS
		ACETATO	METABOLISMO (CORAZON)
		PALMITATO	METABOLISMO (CORAZON)
		METILALBUMINA	HEMATOCRITO DE TEJIDOS (RELACION DE CELULAS A PLASMA)
		OCTILAMINA	DISTRIBUCION DE RECEPTORES DE DE MONOAMINOOXIDASA (PULMON)
		GLUCOSA	METABOLISMO
		2-DESOXI-GLUCOSA	METABOLISMO
		FENITOINA	NIVEL DE DROGAS EN TEJIDOS (EPILEPSIA)
		TIMIDINA	METABOLISMO (TUMORES)
		DOPAMINA, NOREPINEFRINA	FARMACOLOGIA DE LOS NEUROTRANSMISORES (CEREBRO)
		FLUNITRAZEPAM, ETORFINA, PIMOZIDE	FARMACOLOGIA DE LOS NEURORRECEPTORES (CEREBRO)
FLUOR 18	~110 MINUTOS	2-DESOXI-D-GLUCOSA	METABOLISMO
		3-DESOXI-D-GLUCOSA	METABOLISMO
		HALOPERIDOL, SPIROPERIDOL	FARMACOLOGIA DE LOS NEURORRECEPTORES (CEREBRO)

SUSTANCIAS EMISORAS DE POSITRONES, preparadas para exploración mediante TEP. Encuentran aplicación en una amplia variedad de procesos biológicos. Los isótopos, de corta vida, del oxígeno, nitrógeno y carbono son de particular utilidad para seguir el metabolismo de estos elementos, siendo también posible incorporarlos a compuestos de interés farmacológico. El isótopo de larga vida flúor 18 sustituye al oxígeno en análogos de metabolitos propios del organismo y en productos farmacéuticos.

En particular, la medición del flujo sanguíneo en el cerebro ha trascendido las posibilidades de la mayoría de los sistemas de TEP en funcionamiento, al ser estos incapaces, por demasiado lentos, de tomar los datos necesarios para la reconstrucción adecuada de un corte de tejido en el tiempo disponible, que no pasa de unos segundos. Por esta razón se han ensayado métodos alternativos. Uno de ellos consiste en construir una imagen en equilibrio del agua marcada con

oxígeno 15, la cual se obtiene haciendo que el paciente inhale ^{15}O-anhídrido carbónico, y deducir de ello el flujo sanguíneo. A pesar de que la cantidad de agua, marcada con oxígeno, existente en el cerebro es una clara función del flujo sanguíneo, no está relacionada linealmente con él. Además, por razones no bien comprendidas, el método parece dar resultados desiguales.

De entre los métodos disponibles quizás el más prometedor sea la modificación de otro que se preparó en un principio para medir el flujo sanguíneo en ratones y otros animales de laboratorio. En él, se suministra un trazador marcado y fácilmente difundible por vía intravenosa durante un minuto. Se decapita luego al animal. Se determina la cantidad de trazador existente en cortes reales del cerebro mediante autorradiogramas y se le relaciona con la curva de actividad de trazador-tiempo en la sangre arterial del animal hasta el momento de su muerte. Partiendo de estos dos conjuntos de datos, podemos calcular el flujo sanguíneo. No parece difícil sustituir los autorradiogramas por imágenes obtenidas por TEP y llegar a calcular el flujo sanguíneo en sujetos vivos.

La tecnología de la tomografía de emisión de positrones se halla en un estado avanzado de desarrollo. Los biólogos y, un poco rezagados, los clínicos, van comprendiendo cada vez mejor sus posibilidades y sus limitaciones. Muchos estudios con TEP han obtenido resultados inalcanzables para cualquier otra técnica; entre ellos, la determinación del metabolismo regional en el cerebro y el corazón, estudios de permeabilidad de tejidos, medidas del tamaño de infartos remanentes en el corazón a consecuencia de ataques coronarios y prometedoras investigaciones en la fisiología de las psicosis. El gran futuro médico de la tecnología TEP se halla en su verosímil poder de proporcionar atisbos de situaciones patológicas difíciles o imposibles de conseguir por cualquier otro medio. La TEP se presenta como una herramienta particularmente útil en la determinación de los efectos de productos químicos sobre tejidos enfermos o que sufren disfunciones. Un ejemplo impresionante es la capacidad demostrada por la TEP de revelar cambios metabólicos regionales aparentemente invertidos por farmacoterapia en pacientes esquizofrénicos.

En oncología clínica, la TEP ofrece la posibilidad de considerar tejidos malignos como una entidad metabólica y de evaluar el efecto del tratamiento –tanto si se trata de terapia de radiación, hormonal o química– a partir de los cambios registrados en el tejido maligno y de las reacciones bioquímicas de los tejidos normales envolventes. En la práctica normal, el tamaño del tumor maligno y su respuesta a la terapia se fundan casi por entero en exámenes morfológicos: observación directa, palpación incluida, de tumores accesibles o a través de rayos X cuando éstos son inaccesibles.

El precio de la tecnología de TEP es alto. Requiere disponer de un ciclotrón en el propio centro hospitalario, instalaciones químicas adecuadas y uno o más sistemas de captación de imágenes y equipos de cálculo, gobernado todo ese equipo por un grupo de físicos, químicos, matemáticos, fisiólogos y médicos capaces y dispuestos a trabajar conjuntamente. Se han formado ya grupos que trabajan activamente en unas 40 instituciones dispersas por todo el mundo. En gran medida, el desarrollo de la TEP se debe a la Comisión de Energía Atómica (a la que sucede ahora el Departamento de Energía), que patrocinó algunos de los primeros esfuerzos en este campo, al Instituto Nacional de Corazón y Pulmón, el cual valoró, desde el comienzo, el potencial encerrado en esta técnica para la investigación de enfermedades del cerebro y el corazón, y más recientemente al Instituto Nacional de Trastornos y Ataques Neurológicos y de Comunicación, que tomó la iniciativa de financiar las investigaciones sobre TEP en numerosos centros hospitalarios. (Todas esas entidades son norteamericanas.) Desde un punto de vista comercial, varias empresas están fabricando ciclotrones y sistemas completos de imagen y cálculo para TEP.

La resolución obtenida ahora por los instrumentos comerciales más modernos se halla en torno al centímetro para exposiciones de un minuto. Parece razonable esperar que la nueva generación de aparatos mejorará la resolución, en un factor dos, con exposiciones de pocos segundos. La incorporación al proceso de reconstrucción de información sobre el tiempo de propagación del fotón, que exige una estimación exacta de la diferencia de tiempo de llegada de los dos fotones a los detectores de coincidencia, promete aumentar la calidad de las imágenes considerablemente, al mejorar la relación señal/ruido total. La TEP ha alcanzado un nivel en el cual su futuro en el campo de la biología y la medicina parece asegurado.

Minería del petróleo

Con la perforación y el bombeo tradicionales se extrae menos de la mitad del petróleo de un campo corriente. Debido al aumento de los precios de los crudos, su explotación minera reviste un innegable interés económico

Richard A. Dick y Sheldon P. Wimpfen

El sistema tradicional de explotación de un campo petrolífero es notoriamente ineficaz. La cantidad de petróleo que puede extraerse de un yacimiento subterráneo, por bombeo de los pozos perforados, puede ser menos de la décima parte del total y, por término medio, considerablemente inferior a la mitad. En los tiempos de abundancia y bajos precios, este tipo de extracción ineficaz mereció poca atención, pero hoy día, con la continua merma de las reservas petrolíferas interiores y los precios de los crudos importados, los países no deberían olvidar, por más tiempo, aumentar sus disponibilidades. Gran parte del petróleo abandonado bajo tierra después del bombeo puede someterse a explotación minera.

La idea de una minería del petróleo puede parecer extremada. La verdad es que en los Estados Unidos se trata de un hecho casi tan antiguo como el de la propia perforación. Además, el petróleo ha sido, y sigue siendo aún, satisfactoriamente explotado por minería en todo el mundo. En Ohio y en California el sistema de minería se aplicó a la extracción del petróleo hace más de cien años, de dos formas distintas. En Macksburg (Ohio), en 1865, cuando un pozo convencional suministraba poco crudo y mucha agua, se paraba la perforación y se excavaba un pozo de gran diámetro que sustituyera el pequeño orificio inicial. El crudo se separaba luego por decantación de la superficie del agua. Desgraciadamente, el nuevo pozo apenas si rendía poco más crudo del que se obtenía antes.

Al año siguiente, Josiah Stanford, un pionero de la industria del petróleo en el Oeste, atraído por los afloramientos del petróleo existentes en la Sulphur Mountain, al norte de Ventura (California), excavó una serie de túneles en la ladera de la montaña, ligeramente inclinados hacia arriba. Antes de Stanford se habían construido 30 túneles, cada uno de los cuales suministraba entre uno y 20 barriles de crudo por día, que se recogía por gravedad en la boca de los túneles. Estos túneles y un número igual de otros más profundos, excavados en la reserva petrolífera de la Sulphur Mountain a principios de los años 1890, siguieron produciendo petróleo durante varias décadas.

La minería del petróleo en Europa es más antigua todavía. En 1735, en Merkwiller-Péchelbronn, en el departamento del Bajo Rhin (Francia oriental), se excavaron por primera vez unas galerías de mina en el interior de una formación petrolífera. En un comienzo, el petróleo se recogía en zanjas hechas en el suelo de los túneles, pero en el siglo XIX se complementó la captación por gravedad a base de perforaciones. Por los años 30, la red de túneles se extendía ya, subterráneamente, sobre un área de más de tres kilómetros cuadrados. En Wietze, entre Hamburgo y Hannover, en la República Federal Alemana, parece que se desarrolló, a finales del siglo pasado, una operación similar al sistema de decantación empleado en Ohio. Al crecer la demanda de crudos en Alemania durante la primera guerra mundial, se intensificó la producción de Wietze. En ese sentido, se construyó, en 1917, un pozo de cuatro metros de diámetro hasta una profundidad de 151 metros; un año más tarde, se decantaron en dicho pozo 127.000 barriles de crudo.

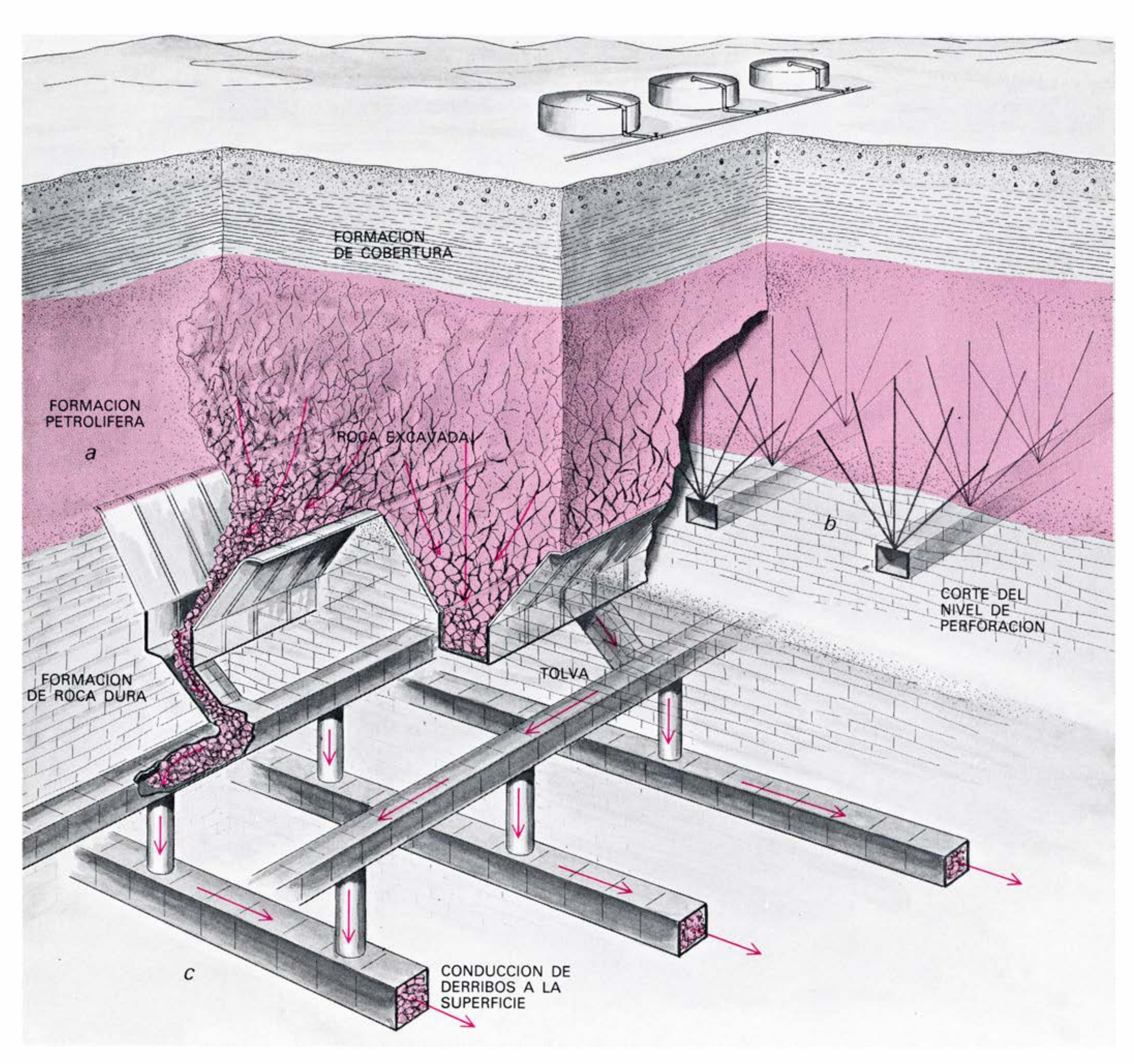
Hoy en día, los mayores esfuerzos en el dominio de la minería mundial del petróleo se registran en Canadá y en la Unión Soviética. El empeño canadiense en concreto se centra en la explotación de las arenas asfálticas ("tar sands"), y no de un yacimiento propiamente dicho; ahora bien, puesto que el producto final es también un crudo, las operaciones resultan ilustrativas. Las arenas asfálticas (areniscas impregnadas de un betún blando) se hallan entre un lecho de rocas calizas y un horizonte de tierra sobrepuesta. El material se explota por el sistema de mina a pozo abierto. Se retiran las tierras sobrepuestas en ringleras de 15 metros de ancho para descubrir los estratos bituminosos, que tienen una profundidad media de 40 metros. La roca blanda se amontona en largas hileras y se transfiere mecánicamente a cintas transportadoras, que la conducen a la planta de procesado. Para producir 45.000 barriles de crudo de petróleo sintético por día, a partir del betún contenido en la roca, es necesario suministrar 117.000 toneladas de arenas asfálticas a la planta de procesado. (El crudo se denomina sintético porque se obtiene por destilación del betún.) Se ha superado ya la meta de los 45.000 barriles diarios de producción, y una segunda instalación, financiada por un consorcio internacional de compañías petrolíferas, deberá producir, próximamente, otros 125.000 barriles por día.

El esfuerzo soviético se ha desarrollado a raíz de los bajos rendimientos obtenidos por perforación en el campo petrolífero de Yarega, en los Urales occidentales, cerca de Ukhta en la República Autónoma de Komi. La reserva del campo, que se encuentra a unos 200 metros de profundidad, consiste en un "aceite pesado", que resulta demasiado viscoso para extraerlo por perforación y bombeo convencionales. Se estimó que, por bombeo desde la superficie, se había recuperado sólo un dos por ciento de las reservas. Para mejorar la situación se perforaron pozos de gran diámetro hasta una profundidad de más de 150 metros y, a partir de cada uno de ellos, se excavó una red de galerías horizontales. Desde estas zonas de trabajo se construyeron pasadizos inclinados hacia el interior de la formación petrolífera, excavándose una galería de producción, de 25 metros de diámetro, en el fondo de cada

pasadizo. Después, se efectuaron una serie de perforaciones radiales en cada galería de producción en sentido horizontal y ligeramente inclinadas hacia arriba, hasta un radio de más de 245 metros.

Con el fin de facilitar el flujo del aceite pesado hacia estos orificios colectores, se inyectó vapor de agua en la roca almacén, a través de orificios verticales practicados desde las galerías horizontales superiores. El crudo, ya algo caliente, se calentaba más en una terminal central subterránea antes de ser bombeado a la superficie. Los rusos estiman que la combinación de perforaciones laterales desde las galerías de producción subterráneas y la inyección de vapor permitirá recuperar del 50 al 60 por ciento de las reservas de petróleo del campo de Yarega, a un ritmo de más de seis millones de barriles por año. Mientras tanto, han iniciado una operación minera similar en un campo agotado al norte de los pozos de Baku, en la República de Azerbaidjan, considerando la posibilidad de hacer lo mismo en más de 300 bolsas de aceites pesados en otros puntos de la Unión Soviética.

¿Qué cantidad de petróleo puede llegar a obtenerse mediante explotación minera en los Estados Unidos? Se ha estimado que entre el volumen que en la actualidad resta en sus campos petrolíferos agotados y el que restará en los campos todavía productivos o en reserva suman un total de más de 300.000 millones de barriles de petróleo. La magnitud de estas reservas potenciales puede apreciarse mejor comparándola con las reservas interiores supuestamente recuperables por métodos convencionales: menos de 30.000 millones de barriles. Desde luego, no todos los 300.000 millones de barriles que restan en los campos norteamericanos pueden recuperarse me-

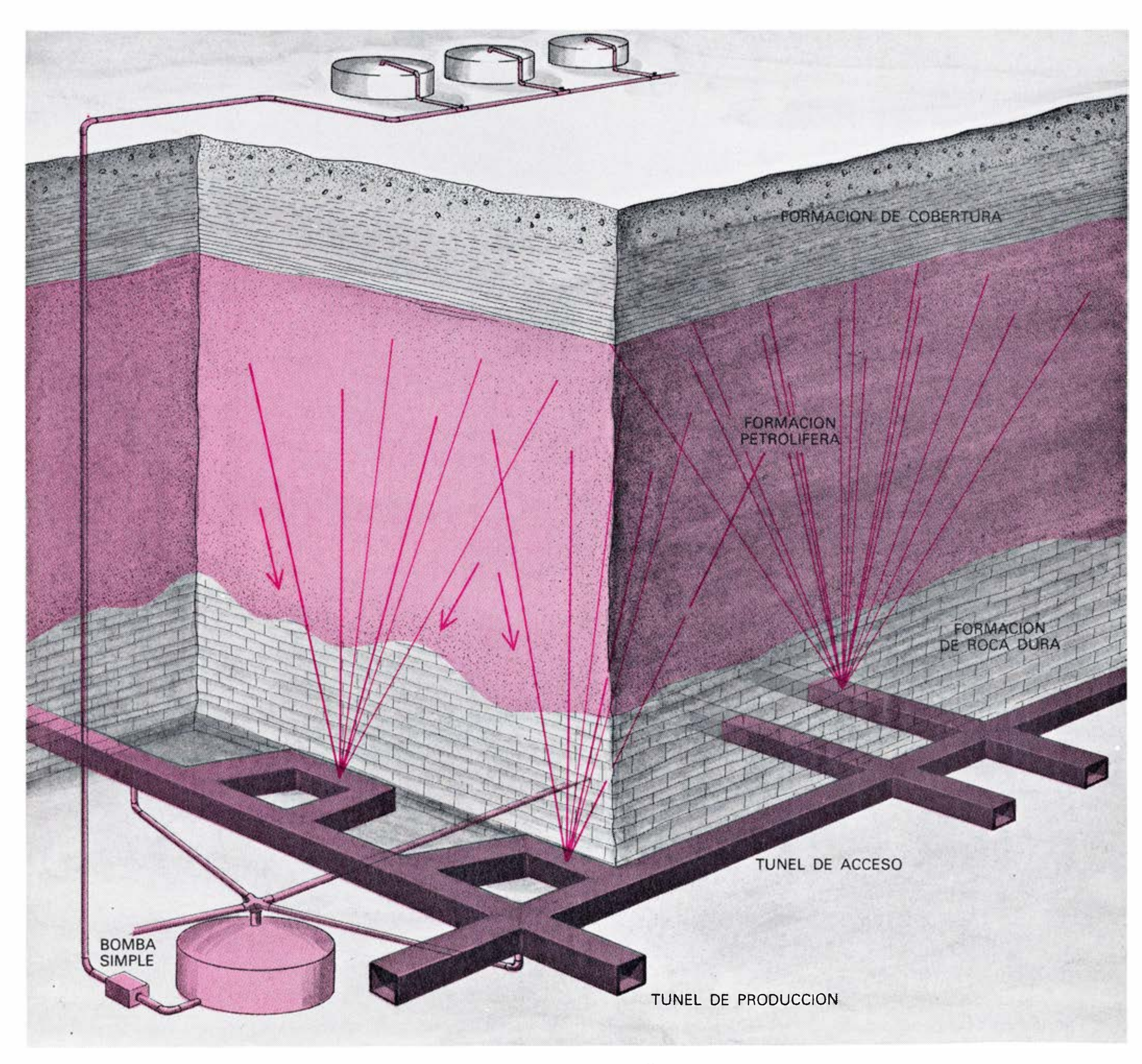


EL PETROLEO DE ROCAS BLANDAS podría extraerse mediante minería subterránea. El primer paso consistiría en la excavación de una serie de túneles conectados entre sí, en una formación de rocas duras, subyacente a la formación petrolífera. La parrilla de túneles superiores comunicaría con el fondo del yacimiento petrolífero (*a*) a través de unos conductos (tolvas) ascendentes. La perforación, voladura y excavación por debajo (*b*) produciría el derrumbe de la roca blanda. Después, sus restos pasarían, a través de las tolvas de los túneles superiores, a una segunda red de túneles (*c*) para su transporte a la superficie. Allí, una planta de aprovechamiento separaría el crudo de los restos de roca. El método se llama de laboreo por socavación.

diante minas. Pero la cifra de 300.000 millones de barriles tampoco tiene en cuenta otras importantes fuentes de hidrocarburos. Si a las reservas estadounidenses de aceites pesados similares al crudo de Yarega se le añaden ciertos depósitos diatomáceos impregnados de hidrocarburos de California y los esquistos bituminosos de Utah (los cuales se cree que contienen unos 30.000 millones de barriles de crudo sintético), resulta que virtualmente se dispone de otros 200.000 millones de barriles adicionales. En esta cifra se excluye, aún, la producción de petróleo sintético a partir de los esquistos bituminosos del Green River, en Utah, Colorado y Wyoming. Esta formación posee unas reservas estimadas de 1,8 billones de barriles, de los que parecen recuperables entre 400.000 y 600.000 millones de barriles. Pero aquí no vamos a ocuparnos de la minería de los esquistos bituminosos.

No hace mucho, el Bureau of Mines, un servicio adscrito al Departamento de Interior de los Estados Unidos encargó dos estudios con el fin de establecer una evaluación actualizada del potencial de la minería del petróleo norteamericana. El último de tales estudios lo había efectuado el propio Bureau hace más de 40 años, cuando el petróleo producido por métodos convencionales costaba más de dos dólares por barril y las reservas norteamericanas parecían inagotables. Para presentar los resultados de los recientes estudios en su contexto adecuado, debemos primero describir ciertos aspectos de la tecnología minera y después discutir tres criterios amplios de posibilidades reales.

Los métodos de extracción de minerales pueden clasificarse en tres apartados: minería de superficie, minería in situ y minería subterránea. Las operaciones con las arenas asfálticas en Alberta son típicas de la minería de superficie; la perforación convencional (no sólo para el petróleo sino también para el azufre y otros minerales solubles) constituye un ejemplo de minería in situ. Aunque la minería subterránea puede presentar varias formas, desde el punto de vista de la minería del petró-

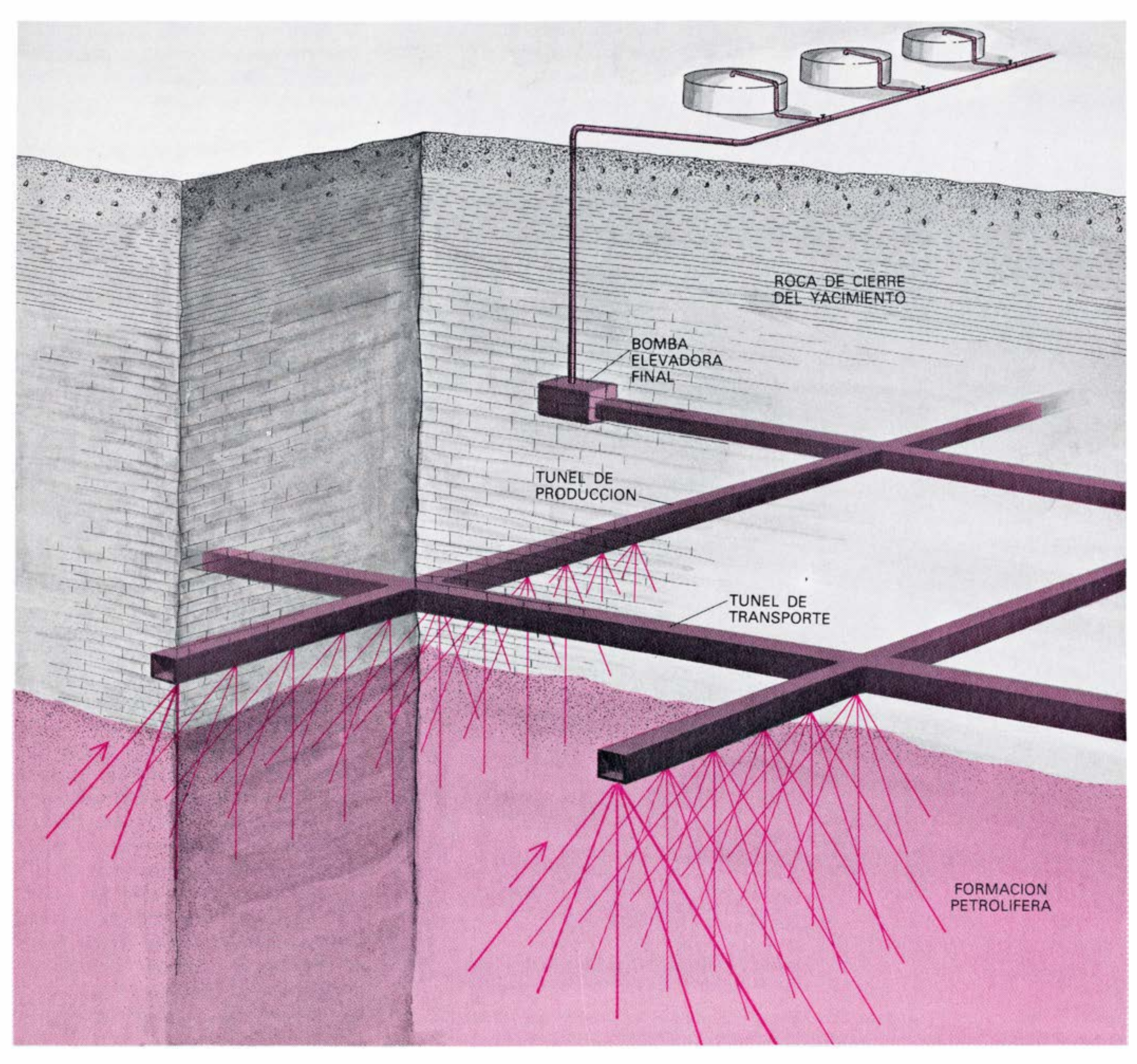


SE LLAMA DRENAJE POR GRAVEDAD al método de minería del petróleo que también depende de la existencia de una formación rocosa dura por debajo del yacimiento petrolífero. Se excava una red de túneles (*a*) en la roca dura, lo más cerca posible a la base del yacimiento. Se perfora un gran número de orificios colectores hacia el interior de la formación petrolífera. El petróleo desciende por gravedad a través del entramado de tubos hasta un tanque colector y se bombea luego a la superficie para su refino. El sistema simple de bombeo hace que este método sea económicamente rentable.

leo los métodos adecuados son limitados. Por ejemplo, no resulta práctico la perforación bajo tierra y el transporte de la roca petrolífera hasta la superficie, tal como se hace con el carbón bituminoso. Las acumulaciones petrolíferas que existen en formaciones rocosas suficientemente resistentes para que sea viable la construcción de túneles poseen una baja proporción de aceite respecto a roca. Por contra, cuando la proporción de aceite es lo bastante elevada como para hacer económicamente atractiva la minería, es la propia roca la que posee una fragilidad peligrosa.

Hay un sistema más prometedor que combina los métodos subterráneo e in situ, como en las minas rusas de petróleo en Yarega. Requisito previo para estas operaciones in situ "modificadas" es la existencia de estratos rocosos, estructuralmente fuertes, en las proximidades de la acumulación petrolífera, y, con preferencia, bajo la acumulación, pero no más de 30 metros por debajo de ella. La primera etapa de una tal operación consistiría en la construcción de un pozo hasta alcanzar la roca fuerte situada bajo la roca-almacén; desde el fondo de este pozo se extenderían una serie de galerías horizontales formando un entramado de calles, que cubrirían toda la zona productiva.

En orden ascendente, encontramos en un campo petrolífero típico agua, aceite y gas. Al trabajar en los túneles subyacentes, los mineros deberían perforar hacia arriba, hasta alcanzar el crudo; el correspondiente entubado, que aislaría la zona del agua, permitiría, una vez perforado en la zona del crudo, el drenaje por acción de la gravedad del aceite hacia un sistema colector situado en los túneles inferiores. Desde allí, el aceite podría bombearse por el pozo de la mina hasta la superficie. Teniendo en cuenta que los agujeros perforados deberían ser de escasa longitud, no más de 30 o 45 metros, se sacaría mayor provecho económico abriendo muchos orificios próximos entre sí, para promover un máximo drenaje. Con controles apropiados, el yacimiento se iría vaciando de un modo uniforme, y así se obten-



SE LLAMA DRENAJE POR BOMBEO al método alternativo de explotación de yacimientos petrolíferos no totalmente agotados y que tampoco yacen sobre formaciones rocosas consistentes. Los túneles de acceso se excavan por encima del yacimiento y se practica un gran número de perforaciones descendentes, hacia el interior de la formación petrolífera. El método es menos económico que el del drenaje por gravedad, porque cada conducto precisa de su propia bomba. En la URSS se recuperan, por inyección de vapor, crudos no explotables mediante bombeo superficial convencional.

dría un elevado porcentaje de recuperación del aceite. Como en Yarega, se podría utilizar, en caso necesario, la estimulación mediante vapor, para asegurar la fluidez del aceite a través de los orificios perforados.

En uno de los estudios encargado por nosotros, que tomó el campo de Wheeler Ridge, cerca de Barkersfield (California), como modelo, se estima que el coste medio del petróleo obtenido con la combinación de minería subterránea y drenaje por gravedad sería de 14 dólares por barril. (El coste, calculado en dólares 1977, incluye una asignación para regalías [royalties] y beneficios).

En la exploración petrolífera, sin embargo, es costumbre detener la perforación una vez que el sondeo ha entrado en la formación petrolífera. Nadie sabe, pues, cuántos campos petrolíferos hay en los Estados Unidos con estratos rocosos subyacentes, suficientemente consolidados como para permitir la excavación de los túneles que las minas de gravedad requerirían. Antes de realizar ningún anteproyecto sobre tales minas, convendría llevar a cabo un ambicioso programa de sondeos exploratorios, para determinar exactamente qué campos petrolíferos sobreyacen a estratos rocosos fuertes y suficientemente próximos a su base.

A la inversa, las prácticas normales de exploración del petróleo proporcionan ya un registro de muchos yacimientos americanos cubiertos por estratos rocosos consolidados, susceptibles de aceptar un sistema de extracción in situ modificado, parecido al empleado en Yarega. En este caso, el sistema de túneles debería excavarse por encima del yacimiento y las perforaciones tendrían que dirigirse hacia abajo, hasta penetrar en él. En lugar del drenaje por gravedad, cada conducto perforado debería ser equipado con una bomba. Esta combinación de minería y bombeo resultaría necesariamente más compleja y cara que el sistema simple de drenaje por gravedad, pero no como para hacerse prohibitiva. En tal sentido, y en otro de nuestros estudios en que se tomó por modelo el yacimiento de Irma, en Arkansas meridional, se estimaba el coste del crudo producido en este tipo de mina de bombeo en 16 dólares por barril. (Esta estimación y todas las que se citarán a continuación se basan también en dólares de 1977.) Si se precisa inyección de vapor, como en Yarega, el costo asciende entonces a 24 dólares por barril. Con los precios de los crudos importados, por encima de los 30 dólares por barril, en 1980, tales minas de bombeo empiezan a adquirir interés económico.

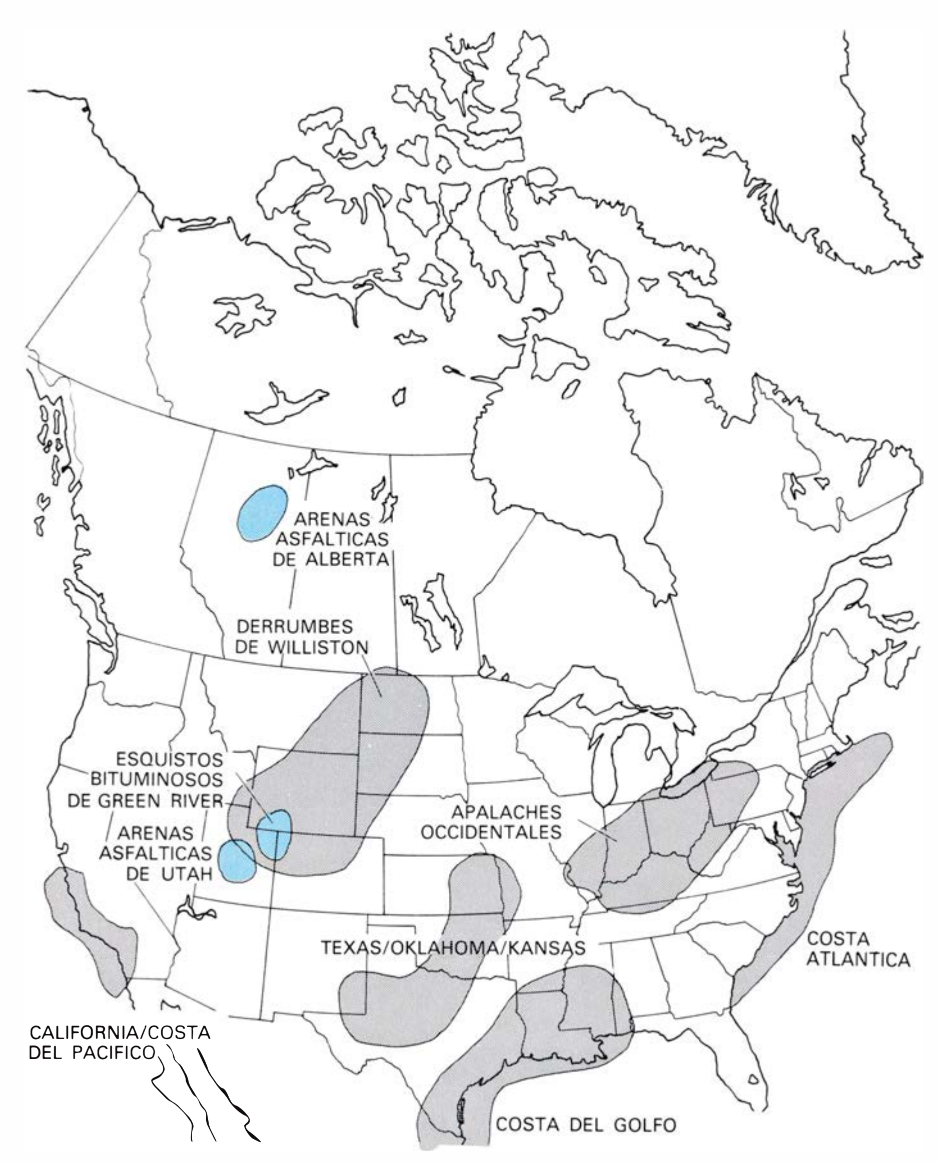


ARENAS ASFALTICAS DE ALBERTA; se explotan por medio de técnicas de minería de superficie. Se aprecian dos de las zanjas excavadas en un depósito de arenas asfálticas por la Syncrude Canada Ltd. Las excavadoras amontonan la roca asfáltica blanda en ringleras; se pueden ver dos cerca del final de las zanjas. La roca se transfiere entonces a las cintas transportadoras por medio de ruedas de cangilones; puede observarse una justo encima de la excavadora de la izquierda. Las largas cintas transportadoras depositan la roca en la planta de procesado, donde el asfalto se convierte en petróleo sintético.

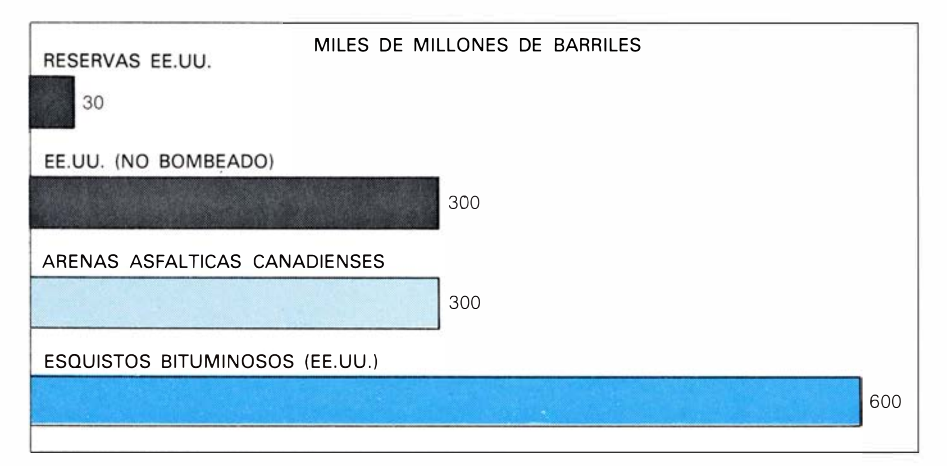
Las consideraciones económicas, aunque de primordial importancia, constituyen sólo uno de los tres criterios básicos que conciernen a la viabilidad de la minería del petróleo. Los otros dos son el grado de seguridad con que trabajan los mineros y el alcance de las perturbaciones del ambiente. Ningún sistema de minería es aceptable si perjudica indebidamente la seguridad o la salud de los mineros o degrada el ambiente, o ambos a la vez. (El término indebidamente es apropiado, en este caso, porque no existe ningún tipo de minería que no exponga a los mineros a algún riesgo y no altere, en alguna forma, el ambiente.) En estos dos aspectos, los sistemas in situ modificados se catalogan desde satisfactorios hasta excelentes.

Con respecto a la seguridad laboral, cualquier trabajo subterráneo que se efectúe en las proximidades de depósitos de hidrocarburos comporta los riesgos asociados con los fuegos y las explosiones. En los sistemas in situ modificados, por ejemplo, la red de túneles se encontraría a cierta distancia del yacimiento propiamente dicho y protegida por rocas fuertes. Además, el crudo extraído del yacimiento viajaría hasta la superficie en un sistema cerrado. La posibilidad de fugas de gas tampoco debería existir. Esos lugares de trabajo requerirían, ciertamente, normas de seguridad tan estrictas como las impuestas en las minas subterráneas de carbón.

Por lo que respecta al impacto ambiental, la minería subterránea presenta un considerable potencial de degra-



LAS CINCO REGIONES PRINCIPALES productoras de petróleo de los Estados Unidos aparecen junto con la todavía zona improductiva del Atlántico. También se muestran (*color*) tres fuentes no convencionales de hidrocarburos: las arenas asfálticas de Alberta y Utah y los esquistos bituminosos de Green River, en Utah, Colorado y Wyoming. Cada una de estas regiones posee yacimientos agotados, donde el petróleo puede extraerse por minería. El más prometedor está en California; allí, además de los yacimientos de petróleo, hay un extenso lecho de sedimentos diatomáceos que contiene hidrocarburos.



RESERVAS ESTIMADAS DE PETROLEO EN LOS ESTADOS UNIDOS (unos 30.000 millones de barriles). Dichas reservas se comparan con las supuestas cantidades que todavía quedan o quedarán en los yacimientos agotados (300.000 millones de barriles). Estas son iguales que la cantidad total de petróleo sintético disponible en las arenas asfálticas de Alberta (*color claro*) y alrededor de la mitad de la recuperable a partir de los esquistos bituminosos de Green River (en los estados contiguos de Utah, Colorado y Wyoming), que pueden suministrar más de 57 litros de aceite por tonelada (*color oscuro*).

dación, en razón de los residuos mineros y la subsidencia geológica (el hundimiento del suelo una vez minado). El sistema in situ modificado para la minería del petróleo, en cambio, constituye una amenaza mínima para el ambiente, desde ambos puntos de vista. La cantidad de material excavado y sacado a la superficie para su vertido posterior es pequeña, en comparación con la que proporciona la extracción minera convencional; las perspectivas de hundimiento son también escasas. La mayor parte del material extraído de la mina es el propio crudo. Las cantidades de roca y agua producidas deberían ser pequeñas, y esta última podría reinyectarse en el yacimiento, para minimizar posteriormente la probabilidad de hundimiento. Resumiendo, la minería del petróleo in situ modificada resultaría un proceso notablemente limpio.

En los Estados Unidos, la mayor parte de la extracción minera no se realiza subterráneamente, sino por alguna de las tres técnicas de minería de superficie: la minería de desmonte, a pozo abierto y en terrazas. Estas mismas técnicas son aplicables a la minería del petróleo y las tres poseen una característica en común. Todas exigen el movimiento de tierras y rocas en gran escala y, por consiguiente, de grandes inversiones en maquinaria pesada, si se quieren obtener unidades de costo que sean económicas.

Así como los sistemas in situ modificados resultan apropiados para la explotación de yacimientos de 150 o más metros de profundidad, las técnicas de minería de superficie son las idóneas para depósitos menos hondos. Vamos a describir los tres métodos uno por uno. La minería de desmonte ("strip mining"), una técnica frecuentemente utilizada en los campos de carbón del Medio Oeste y del Oeste norteamericanos, empieza con la remoción de las tierras sobrepuestas excavando una larga y estrecha zanja, con una draga de arrastre o una pala excavadora ancha. La zanja suele medir unos 15-30 metros de ancho por más de 1,5 kilómetros de largo; es lo suficientemente profunda como para dejar el mineral a la vista. Se recoge luego el mineral y se transporta hasta la planta de procesado. Se excava una segunda zanja paralela a la primera. Las tierras removidas de esta zanja se vierten sobre la primera y se extrae el mineral. Y así, sucesivamente.

Este sistema es eficaz para depósitos de hasta 55 metros de profundidad, generalmente en terreno llano. La res-

tauración de estos terrenos comporta una serie de procesos relativamente simples: allanamiento de las tierras sobrepuestas, sustitución de la capa de suelo superior y sembrado de nuevo. En muchos lugares del Medio Oeste, campos que fueron objeto de minería de desmonte se están reconvirtiendo con fines agrícolas.

La minería a pozo abierto ("open-pit mining") resulta adecuada para los depósitos de mineral con topografía superficial irregular y, también, para aquellos que se encuentran a profundidades superiores a los 55 metros. Tanto las tierras superiores como el mineral se retiran con anchas palas excavadoras; suelen acarrearse fuera del pozo a través de un sistema de carreteras. También pueden usarse cintas transportadoras. Las cargas de mineral se llevan hasta la planta procesadora y las tierras sobrepuestas se vierten a cierta distancia del pozo. En general, resulta poco práctico utilizar esas tierras para rellenar el pozo.

La tercera técnica de minería de superficie, la minería de terrazas, es una variante de la anterior; se opta por ella cuando el depósito de mineral cubre una extensa área, pero a escasa profundidad. Puede considerarse como un tipo de minería de desmonte en espiral, con la salvedad de que las tierras sobrepuesas se transportan lejos, almacenándose, al menos temporalmente, en lugar de servir para el rellenado del pozo. Sin embargo, debido a que una mina en terrazas ocupa generalmente una amplia zona de trabajo es práctico utilizar el rellenado.

Los cálculos indican que la producción de petróleo por minería de superficie es económicamente rentable en los Estados Unidos. En un estudio en que se tomó por modelo el yacimiento del Kern River, en California, y la minería de terraza como técnica de explotación, se sugiere que los costos totales del crudo serían de 11 dólares por barril. Otro estudio, basado en la técnica de pozo abierto, proyecta el coste a 15 dólares por barril, con el yacimiento de Edna (California) por modelo. Con el yacimiento de Santa Cruz, también en el estado de California, como modelo y la técnica de desmonte, el coste proyectado es de 18 dólares por barril. Cuando se planeó un sistema modificado de minería en terrazas para el depósito de arenas asfálticas de Sunnyside (Utah) el coste estimado alcanzó los 21 dólares por barril, que no sobrepasa mucho el coste del petróleo sintético obtenido de las arenas asfálticas canadienses.

Aunque la minería de superficie es algo más segura que la in situ modificada, su impacto ambiental es superior. Cerca de las minas pueden aparecer gas, polvo y vapores molestos. Las tierras sobrepuestas y los restos de las plantas de procesado ponen sobre el tapete, por otro lado, un importante problema de almacenamiento. Desde luego es posible, a expensas del precio, reducir o incluso eliminar estas consecuencias negativas de la minería de superficie, pero la cuestión de las restricciones ambientales que van a pesar sobre ella permanece, de momento, sin respuesta. El Bureau of Mines está llevando a cabo un estudio para identificar y cuantificar dichos problemas ambientales y sugerir posibles soluciones a los mismos.

¿Por qué no se explota regularmente en los Estados Unidos el petróleo por minería? Una razón que tuvo su validez en el pasado era que el país podía pasar sin ella. Pero a medida que el precio del petróleo importado ha ido subiendo, esta respuesta va perdiendo justificación. Se han iniciado algunos ensayos a escala piloto, incluyendo la extracción de esquistos bituminosos. Un consorcio de 17 firmas, que incluye varias e importantes compañías petrolíferas de los Estados Unidos, concluyó recientemente una prueba de 30 meses sobre el rendimiento de la obtención de petróleo sintético a partir de los esquistos bituminosos de una formación cercana a Rifle (Colorado). Al término del período de prueba, a finales de 1978, las retortas habían producido casi 100.000 barriles de crudo, a un ritmo de 95 litros por tonelada de esquisto bruto. En otro lugar de Colorado se está investigando en esquistos bituminosos triturados mediante cargas explosivas.

Por su parte, un industrial del petróleo está experimentando con la obtención de hidrocarburos en una mina a pozo abierto, en una formación de sedimentos diatomáceos de 122 metros de profundidad, en el campo de McKittrick (California). Otro industrial está trabajando, en cooperación con el departamento norteamericano de Energía, en la investigación de las perspectivas de recuperación de hidrocarburos en una formación de arenas asfálticas en Utah, utilizando métodos de extracción in situ. Parece seguro pronosticar que en el futuro contemplaremos una explotación acelerada del petróleo por minería tanto en los depósitos de hidrocarburos no convencionales como del que aún existe en los yacimientos agotados después del bombeo convencional.

La flotabilidad del nautilo

El animal reduce su peso y gana movilidad mediante la división de su concha en compartimientos y la eliminación de su contenido líquido. Lo consigue a pesar de la fuerte presión hidrostática, que tiende a reintroducir el agua

Peter Ward, Lewis Greenwald y Olive E. Greenwald

Antes del final del Cretácico, hace unos 65 millones de años, la vida en los océanos era muy distinta de la actual. Entre los más importantes de los grandes habitantes de los mares se contaban los cefalópodos tabicados, moluscos que moraban en el compartimiento más externo de una concha dividida en cámaras mediante la secreción, en su interior, de una serie de septos o tabiques. Debieron su predominio al haber desarrollado la capacidad de mantenerse en flotabilidad neutra: una densidad global (o peso dividido por volumen) prácticamente igual a la del medio que los rodeaba. En particular, los cefalópodos tabicados desarrollaron un órgano capaz de extraer el agua de los compartimientos interiores de su concha. Con el advenimiento de esta facultad (y también con la evolución hacia un chorro de agua direccional), estos animales se liberaron de las limitaciones de sus antepasados, que se arrastraban sobre el fondo, y se convirtieron en los primeros grandes carnívoros que nadaban libremente por el mar.

Los cefalópodos tabicados alcanzaron su mayor diversidad durante los períodos Triásico, Jurásico y Cretácico, hace de 225 a 65 millones de años. En la última parte del Cretácico, su número comenzó a disminuir, quizás en respuesta al número creciente de un nuevo tipo de depredador móvil, los modernos peces óseos, que mantienen su flotabilidad neutra de una manera completamente distinta: poseen una vejiga hinchable y una glándula que la puede llenar de gas. Fuera cual fuere la razón, la extinción de los cefalópodos tabicados era casi total a finales del Cretácico. En la actualidad sólo queda el género *Nautilus*. Los animales de este género constituyen una buena ocasión para estudiar el antiguo sistema de flotación, cuya precisión en ese género único es notable: la diferencia entre el peso de un nautilo maduro (hasta 1400 gramos) y el de un volumen igual de agua de mar puede ser del orden de sólo un gramo.

Las especies supervivientes de *Nautilus* se localizan en la vecindad de los arrecifes de coral que rodean las islas del Pacífico occidental tropical. Raramente se las ve cerca de la superficie; en cambio, sí han sido capturadas a profundidades de hasta 600 metros. Puesto que viven por debajo de la zona donde suelen llegar los buceadores, no es mucho lo que se sabe de su comportamiento y ecología. Se han hecho, no obstante, algunas observaciones. En primer lugar, el análisis del estómago de un nautilo revela fragmentos de crustáceos bentónicos (que viven en el fondo). Además, en las raras ocasiones en las que nosotros mismos hemos observado un nautilo en el mar, el animal se hallaba en el fondo o cerca del mismo. Por último, el nautilo es un nadador lento; un escafandrista puede mantenerse fácilmente a su lado. Todo parece indicar que el nautilo se alimenta comiendo lentamente sobre las pendientes del antearrecife.

Las partes blandas del nautilo vivo constan de dos secciones principales, la cabeza (con mayor exactitud, la cabeza-pie) y el cuerpo. La cabeza está recubierta por el capuchón, un tejido resistente y carnoso que le sirve de escudo. El nautilo puede retraerse en el interior de su concha, dejando al descubierto sólo el capuchón.

La peculiaridad más destacada de la cabeza son los tentáculos, que suman más de 90, número muy superior a los que hay en cualquier otro cefalópodo vivo. Cada tentáculo se aloja en una vaina, en cuyo interior puede retraerse y de la que puede sobresalir. La superficie del tentáculo carece de las ventosas que hay en esos apéndices de otros cefalópodos. Se cubre de una sustancia pegajosa que ayuda a la sujeción de la presa.

Los tentáculos rodean un robusto par de piezas mandibulares que recuerdan el pico de un loro grande. Las mandíbulas, muy calcificadas, le permiten romper el exosqueleto de los crustáceos mejor acorazados. A diferencia de la saliva del pulpo, otro cefalópodo, la del nautilo no contiene toxinas que incapaciten a las presas con que se debate. Su única arma ofensiva son las mandíbulas. Bajo los tentáculos y las mandíbulas hay un pliegue de tejido, llamado hiponomo, que se emplea en la locomoción: expulsa un chorro de agua. El hiponomo consta de un par de pliegues musculares que se arrollan entre sí y forman un embudo muy flexible.

Queda otra característica de la cabeza que merece mencionarse. Como otros cefalópodos, el nautilo tiene ojos prominentes, uno a cada lado de la cabeza; pero a diferencia de los ojos de los demás, los del nautilo están poco desarrollados. Carecen, por ejemplo, de cristalino. La luz penetra a través de una minúscula abertura, y, presumiblemente, las imágenes se forman siguiendo el mismo proceso que en una cámara clara. De nuestra propia experiencia directa con estos animales hemos llegado a la conclusión de que los ojos les sirven únicamente para detectar cambios en la intensidad de la luz. Por otra parte, los tentáculos parecen portar células sensibles a la presencia de sustancias químicas. Estas células podrían compensar la carencia de una visión precisa.

La otra sección principal del nautilo vivo, el cuerpo, incluye un gran saco que contiene los sistemas de órganos del animal. Gran parte del espacio está ocupado por el sistema digestivo y el reproductor. El propio saco está envuelto por el manto, una lámina de tejido que segrega la concha. El manto posterior segrega los septos, que dividen la concha en compartimentos. El espacio que existe entre el manto y el

saco está abierto bajo el cuerpo, creando una gran cavidad que comunica con el hiponomo. La cavidad contiene cuatro grandes branquias, en contraposición al par de branquias único que se encuentra en todos los demás cefalópodos. La cavidad recibe asimismo los poros eferentes de los sistemas digestivo y reproductor.

Este género posee dimorfismo sexual. El macho es ligeramente mayor que la hembra, debido a la presencia de un órgano de gran tamaño, el espádice, que durante la cópula introduce un paquete de espermatozoides en la cavidad del manto de la hembra. Las hembras de muchas especies de cefalópodos producen miles de huevos al año; una hembra promedio de nautilo no produce más de 10 grandes huevos.

La concha en la que vive puede describirse de modo aproximado como un cono hueco estrechamente arrollado sobre sí mismo. Está compuesta por capas de aragonito, una forma cristalina del carbonato cálcico, que alternan con capas de una sustancia proteinácea de composición química semejante a la de la uña.

La división interna de la concha en cámaras hace pensar inmediatamente en la posibilidad de que el nautilo pueda cambiar su flotabilidad, cual si se tratara de un submarino. Ya en 1696 Robert Hooke afirmaba que "el animal tiene el poder de llenar o vaciar cada una de [las cámaras] con agua, de modo que baste para equilibrar y ajustar la postura de su nave, o concha, disponiéndola para la navegación o viaje que tenga que realizar; o bien, si tiene que subir, entonces puede vaciar estas cavidades de agua, o llenarlas con aire". No hemos encontrado nunca un nautilo en estratos someros que volviera a llenar de agua ninguna de sus cámaras; sólo puede vaciarlas. Es más, sólo puede vaciar sus cámaras lentamente; el mantenimiento de la flotación constituye un esfuerzo permanente que durará toda la vida. Al tiempo que el animal crece en su concha, construye un septo tras de sí,



NAUTILO FOTOGRAFIADO durante la noche a una profundidad de 9 metros cerca del arrecife de coral que rodea la isla de Nueva Caledonia, en el Pacífico meridional. El animal se cierne casi sin esfuerzo debido a que su peso es prácticamente igual al del agua que desplaza. A pocos metros de donde se tomó la fotografía, el arrecife comienza a descender hasta una profundidad de varias decenas de metros. El nautilo pudo haber migrado hacia arriba durante la noche (su período de actividad), o bien pudo haber permanecido escondido cerca del extremo superior del arrecife durante el día. Debido a que raramente se ven ejemplares vivos, se sabe muy poco sobre su comportamiento. El buceador que aparece a la derecha es Pierre Laboute.

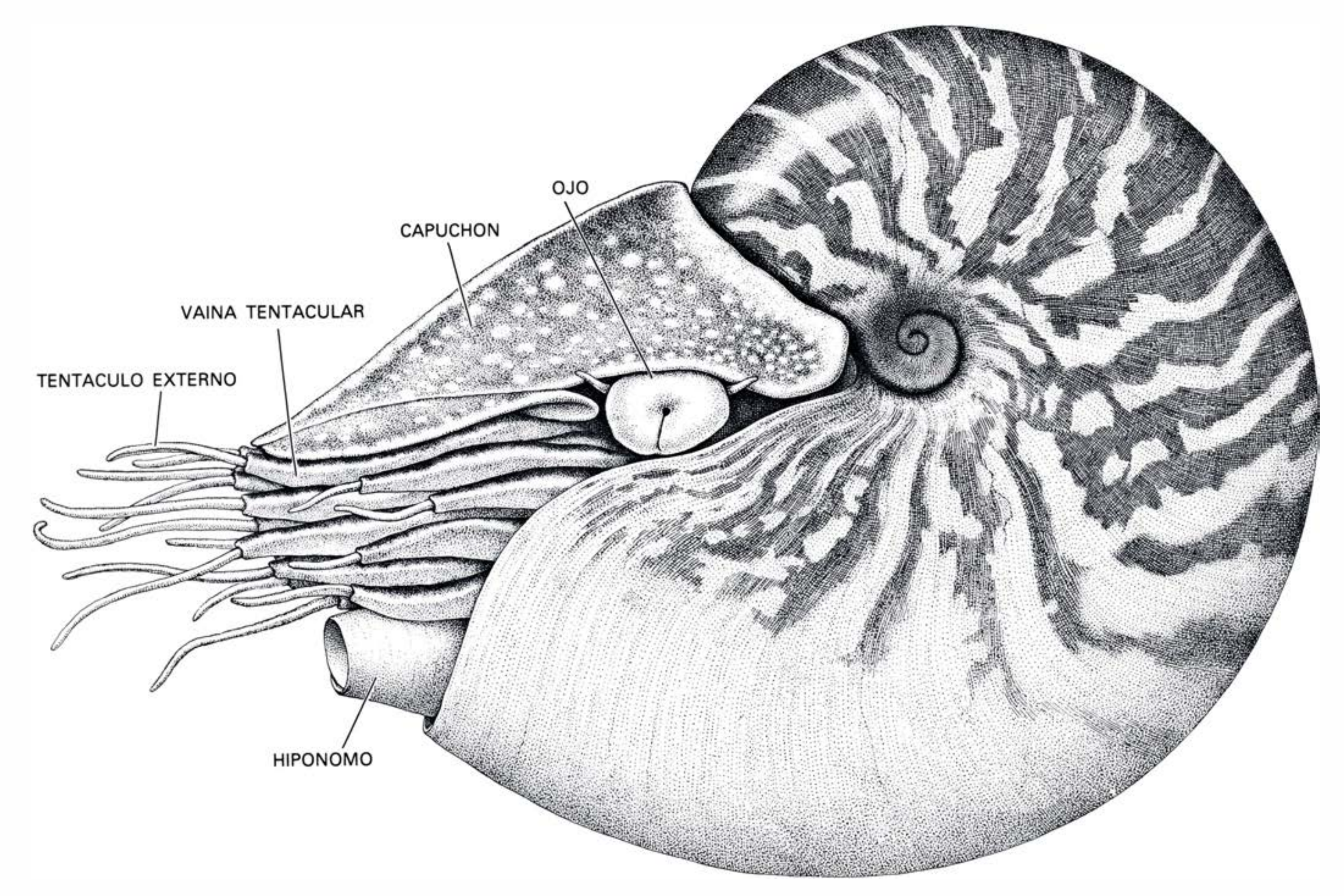
clausurando de este modo una nueva cámara. En principio, cada nueva cámara está llena de un fluido acuoso llamado líquido cameral, pero éste se extrae lentamente. Ello le proporciona el empuje hacia arriba que el animal necesita para contrarrestar el creciente peso de sus partes vivas y de su concha.

La extracción del líquido de las cámaras la realiza el sifúnculo, un filamento de tejido vivo encerrado en un tubo calcáreo que se extiende en espiral desde el manto posterior del animal y pasa por todas las cámaras de la concha, incluidas las más antiguas. En el centro del filamento hay una red de vasos sanguíneos. El agua de mar circundante ejerce una presión hidrostática sobre el cuerpo del nautilo, presión que se transmite a la sangre. De ahí que la presión sanguínea en el sifúnculo sea igual a la presión sanguínea generada por el corazón más la presión del agua de mar, que aumenta en una atmósfera por cada 10 metros de profundidad. A 400 metros, una profundidad típica para el nautilo, el interior del sifúnculo debe tener una presión de más de 40 atmósferas, o más de 40 kilogramos por centímetro cuadrado. Por ello, en una cámara vacía, donde la presión es siempre inferior a una atmósfera, el tubo calcáreo que rodea el tejido del sifúnculo debe impedir que este tejido estalle.

En nuestra investigación nos ocuparemos del mecanismo mediante el cual el sifúnculo extrae el líquido que inicialmente llena cada cámara. El líquido se parece mucho a la sangre del nautilo, y también al agua de mar, por ser las concentraciones en iones sodio, potasio y cloruro prácticamente las mismas en los tres. Por ello resulta imposible afirmar con certeza si el líquido es un filtrado de la sangre, una secreción del sifúnculo o del manto o quizás agua de mar que el animal ha modificado. No obstante, si las células sifunculares transportaran algunos de los iones desde el líquido hasta los vasos sanguíneos del centro del sifúnculo, la concentración iónica del líquido descendería por debajo de la concentración en sangre y el agua del líquido fluiría por ósmosis desde la cámara hasta la sangre, de donde los riñones del animal podrían eliminarla.

Podría objetarse a esta hipótesis que la presión hidrostática en las profundidades del océano favorece la tendencia que muestra el agua de pasar otra vez desde el sifúnculo hasta la cámara. Efectivamente, aun en el caso de que el sifúnculo pudiera reducir a cero la concentración iónica del líquido cameral, y maximizar así la salida osmótica del agua de la cámara, un simple cálculo demuestra que a la profundidad de un poco más de 240 metros la presión hidrostática es lo bastante grande como para conducir el agua en la dirección contraria. Puesto que se encuentran nautilos con las cámaras completamente vacías a profundidades de hasta 600 metros, no dudamos en descartar el mecanismo osmótico simple.

También deseábamos descartarlo experimentalmente. Los ensayos en pleno océano no resultaban viables. Y optamos por abrir un pequeño agujero a través de la concha del nautilo, perforando una cámara casi vacía. Extrajimos todo el líquido que había y añadimos cinco mililitros de una solución cuya concentración iónica era su-



ANATOMIA GENERAL de *Nautilus macromphalus*, la especie estudiada por los autores. La figura de la izquierda muestra las características visibles externamente. De la concha sobresale la cabeza del animal, que tiene más de 90 tentáculos dispuestos en dos espirales sobre cada lado de la boca. Aunque los ojos son prominentes, la visión que permiten no es aguda. La cabeza está protegida por un capuchón carnoso y resistente. La ilustración de la derecha muestra la estructura interna de la concha y del animal vivo. La parte más externa es el manto, un tejido cuyas secreciones amplían la concha. El manto posterior segrega la sucesión de septos, o tabiques, que dividen la concha en cámaras. El sifúnculo, un filamento de tejido vivo encerrado en un tubo

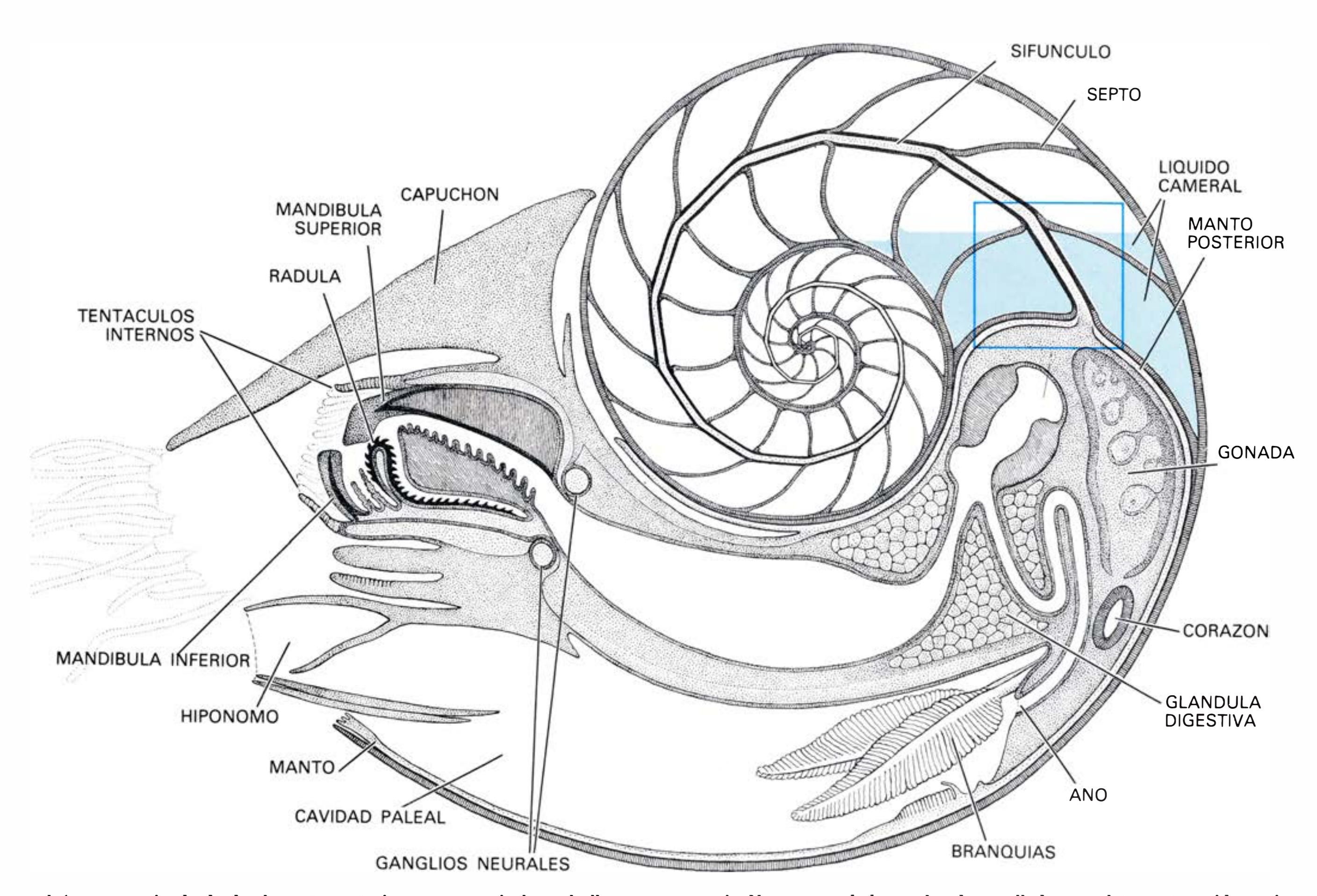
perior a la de la sangre del nautilo. Si el movimiento del agua a través del sifúnculo estaba gobernado por fuerzas osmóticas sencillas, se difundiría agua desde la sangre sifuncular hasta el líquido cameral artificial. Lo que observamos fue lo contrario. El sifúnculo era capaz de extraer el líquido de la cámara aun cuando la concentración de los iones en éste era casi el doble que la de la sangre.

Llegados a este punto, acudimos al trabajo de Jared M. Diamond. En la década de 1960, Diamond y sus colegas de la Universidad de Harvard demostraron que la vesícula biliar del conejo podía transportar agua contra gradientes, fueran osmóticos o hidrostáticos. El agua pasaba desde el lumen (o cavidad) de la vesícula hasta los vasos sanguíneos de la pared de la misma. Los investigadores pusieron entonces de manifiesto que, si bien las células que formaban la pared de la vesícula se hallaban en contacto unas con otras en la superficie dirigida hacia el lumen, había espacios entre las células en los lugares de la pared de la vesícula donde las células se hallaban más cerca de los vasos sanguíneos.

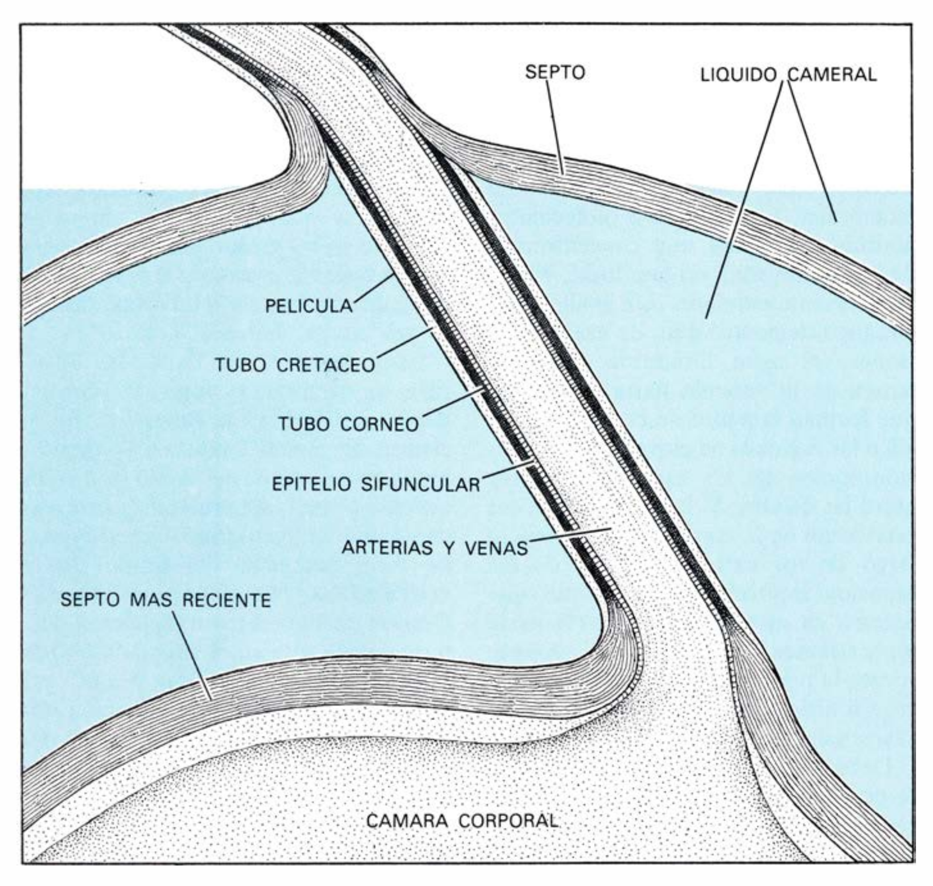
Para los investigadores de Harvard, la membrana celular que rodeaba cada espacio debía incluir enzimas que transportaban iones a través de la membrana. Estas bombas moleculares podrían establecer una concentración de iones elevada, aunque local, en los espacios intercelulares. Un análisis matemático demostró que, en esas condiciones, el agua difundiría desde el lumen de la vesícula hasta las células que forman la pared de la misma, y de allí a las regiones de elevada concentración iónica en los espacios que hay entre las células. Si las bombas iónicas estuvieran en la membrana celular a lo largo de los extremos ciegos de los espacios intercelulares, el agua que entrara en estos espacios fluiría hacia los extremos abiertos de los mismos contra la presión osmótica e hidrostática. En última instancia, entraría en los vasos sanguíneos.

Debido a que esta hipótesis supone ósmosis en pequeñas regiones de una capa celular se denomina modelo de ósmosis local. La importancia del modelo con respecto al nautilo radica en que podría explicar la capacidad del sifúnculo de transportar líquido cameral tanto frente a gradientes osmóticos (como los que producimos en el laboratorio) como contra los gradientes hidrostáticos con los que un animal se enfrenta en las profundidades del océano. Se trata de averiguar si el sifúnculo tiene una geometría celular acorde con el modelo de ósmosis local.

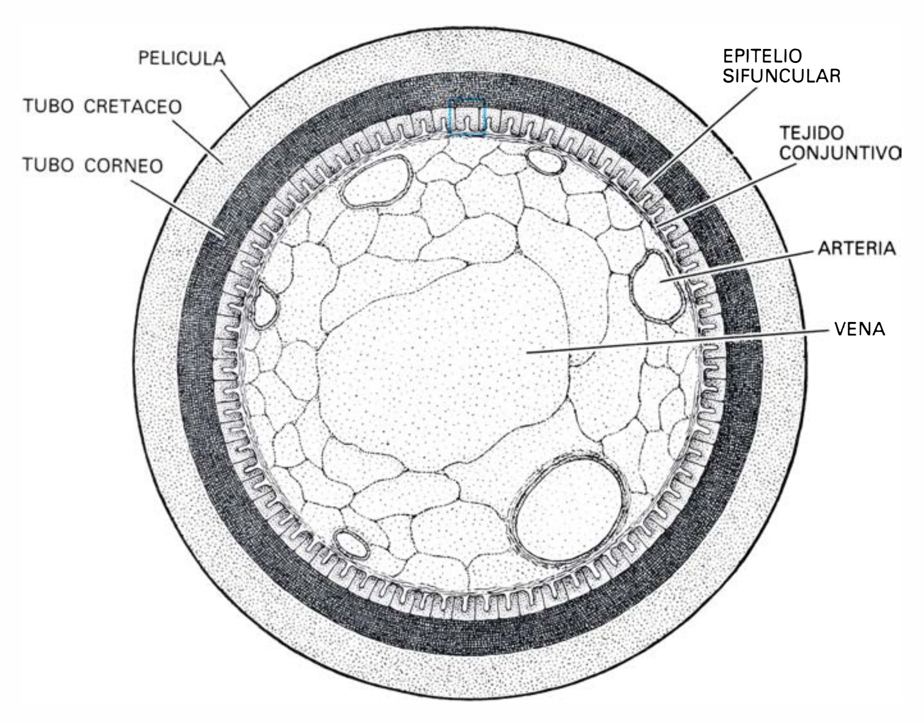
Evidentemente, la tiene. En 1966, Eric J. Denton y John B. Gilpin-Brown, de la Marine Biological Association del Reino Unido en Plymouth, publicaron diseños del tejido sifuncular que mostraban una semejanza sorprendente con las micrografías electrónicas, de baja potencia, del tejido de la vesícula biliar obtenidas por Diamond. Denton y Gilpin-Brown sugirieron posteriormente que en el sifúnculo podría funcionar la ósmosis local. En el examen del tejido sifuncular también nosotros notamos que los segmentos del sifúnculo procedentes de cámaras que se estaban vaciando o que ya se habían vaciado tenían espacios intercelulares conspicuos. Sin embargo, nuestras mi-



calcáreo, se extiende desde el manto posterior en una espiral que lo lleva a través de todas las cámaras. Los dos pares de branquias del animal se hallan extendidos en una cavidad situada entre el manto y las vísceras. La cavidad conecta con el hiponomo, un embudo flexible a través del cual se expulsa el agua en un chorro destinado a la locomoción. Cuando un ejemplar de la especie *N. macromphalus* se ha desarrollado completamente, mide varios centímetros de diámetro y pesa cerca de un kilogramo. En ese estadio, la concha del animal tiene 30 o más cámaras, de las cuales, sólo las más externas contienen líquido. El rectángulo coloreado de la figura de la derecha señala la sección que aparece en la ilustración superior de la página siguiente.



LA CAMARA MAS RECIENTE de la concha es la que se encuentra detrás del animal vivo y cuyo suelo está formado por el septo más reciente. Al principio, la cámara está llena de un fluido acuoso, el llamado líquido cameral, pero éste se va extrayendo lentamente. De este modo, al tiempo que crece, el animal aumenta su flotabilidad para contrarrestar el peso creciente de sus partes vivas y de su concha. Cuando la cámara más reciente empieza a vaciarse, la cámara contigua contiene aún algunos mililitros de líquido.



SECCION TRANSVERSAL DEL SIFUNCULO, el órgano que atraviesa todas las cámaras de la concha siguiendo una disposición espiral y que extrae el líquido de las mismas. Las capas externas del órgano son porosas y calcáreas. Encierran el tejido vivo: una capa epitelial (es decir, una capa de células estrechamente apretadas) en cuyo interior hay una médula compuesta por arterias y venas. La presión de la sangre en el sifúnculo corresponde a la presión del mar alrededor del animal; por ello, en una cámara vacía las capas exteriores del sifúnculo deben impedir que los tejidos internos exploten. El rectángulo de color muestra la parte del sifúnculo que aparece en la ilustración superior de la página siguiente.

crografías electrónicas permitían suponer que los grandes espacios intercelulares podían servir simplemente de canales de drenaje para el fluido procedente de un conjunto de canales mucho más finos que conducían a los espacios mayores. Al proponer que el líquido cameral se transportaba primero hasta el conjunto de canales más finos, nos basábamos en que éstos aparecían tapizados (en el citoplasma que rodea los canales) por numerosas mitocondrias, los orgánulos intracelulares que suministran energía a la célula. Para nosotros, las mitocondrias proporcionan en este caso energía para el transporte de iones o de algún otro soluto.

El tejido procedente de la porción de sifúnculo que atraviesa una cámara cuyo septo todavía se está formando carece de espacios intercelulares, grandes o pequeños. Por otro lado, la parte de sifúnculo de las cámaras ya vacías mantiene su red de espacios. Por tanto, el sifúnculo parece capaz de achicar estas cámaras, que de otro modo se llenarían con el agua forzada hacia dentro por la notable presión hidrostática que existe a profundidades de más de 240 metros.

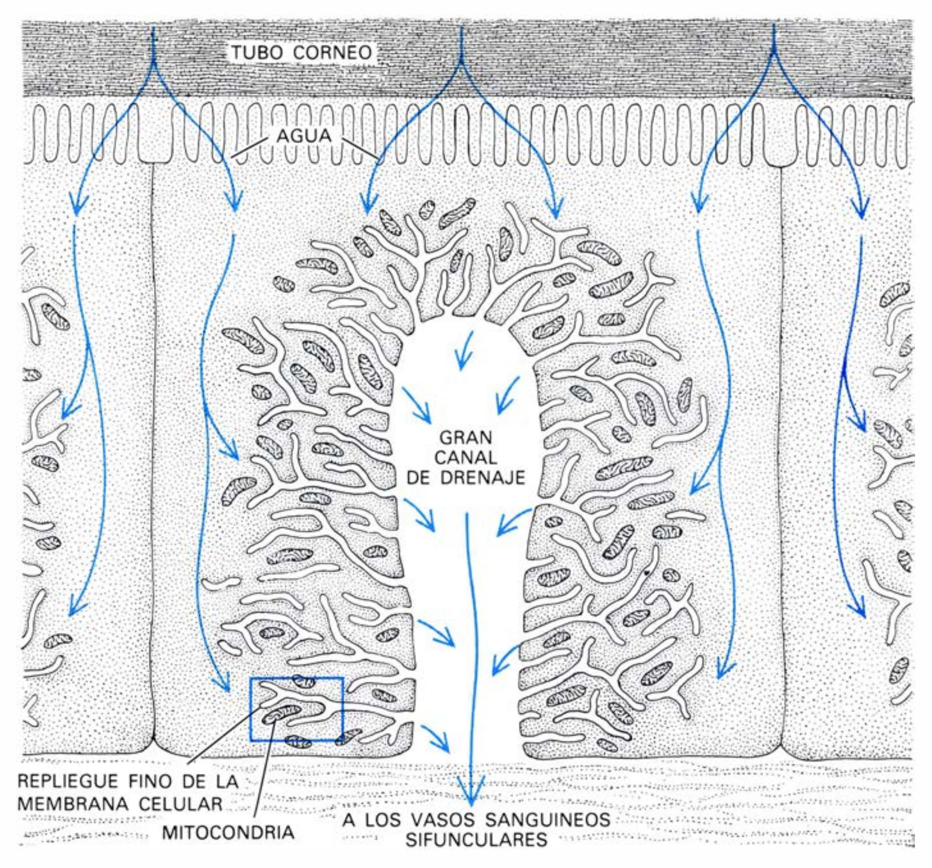
En las cámaras vacías se ha sustituido el líquido por un gas, aire en buena parte, del que se ha eliminado casi todo el oxígeno y al que se ha añadido un poco de dióxido de carbono adicional, presumiblemente secretado por las células sifunculares. Denton y Gilpin-Brown observaron que, cuando una cámara se vaciaba, la presión del gas en su interior podría ser tan sólo de 0,1 atmósferas. Ello les llevó a postular que el gas no se transportaba a la cámara por un proceso que requiriera energía. Quizá simplemente abandonaba por difusión la sangre sifuncular.

Debido a que el agua de mar por donde se desplaza el nautilo se halla en equilibrio químico con los gases atmosféricos de la superficie del océano, y puesto que la sangre del nautilo se mantiene en equilibrio químico con el agua de mar, no resulta sorprendente que las presiones de nitrógeno y de argón de las cámaras llenas de gas sean las mismas que las que estos gases tienen en el aire a nivel del mar. En todo caso, el gas no contribuye a la flotación del animal. En realidad, la reduce ligeramente, porque, después de todo, el gas pesa, en tanto que el vacío no pesa nada. En un globo se introduce gas para establecer su tamaño y su forma. Estas funciones las realiza en el nautilo su propia concha rígida.

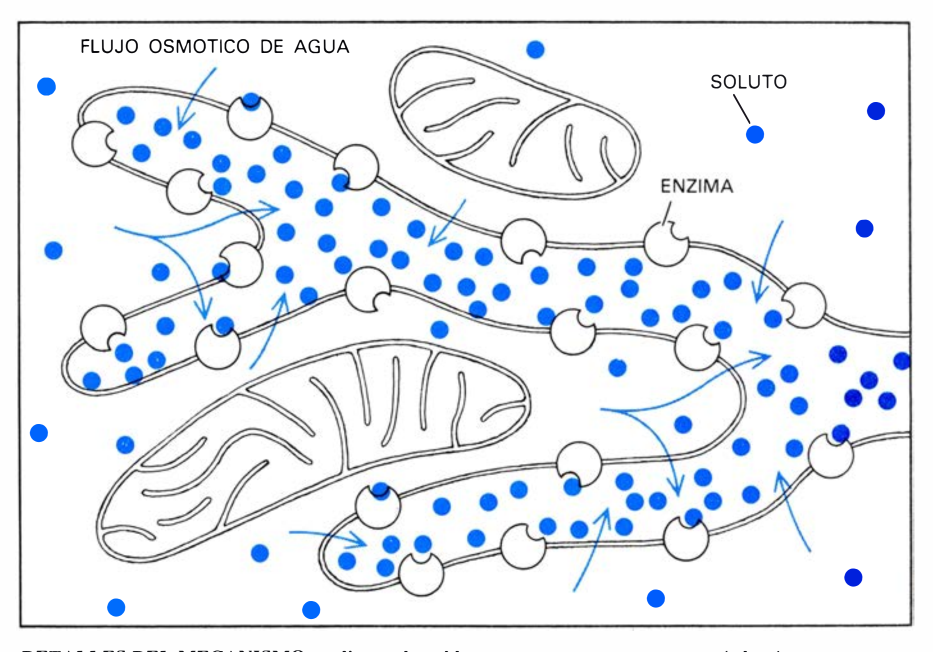
Sin que importe su tamaño, los nautilos pesan de uno a cinco gramos en el mar. El animal no pierde pues su flotabilidad neutra a lo largo de su crecimiento. Ello lo consigue manteniendo en equilibrio dos tasas o velocidades: la velocidad a la que su concha y sus partes blandas en crecimiento aumentan de peso y la velocidad a la que vacía de líquido las cámaras de la concha. Para comprender cómo se hallan relacionadas dichas tasas, Desmond H. Collins, del Royal Ontario Museum, G. E. G. Westermann, de la Universidad McMaster, Arthur W. Martin, Jr., de la Universidad de Washington, y nosotros, abordamos, en distintas investigaciones, varios aspectos del crecimiento de la concha de este cefalópodo.

Las radiografías de un nautilo en desarrollo muestran que el crecimiento abarca una secuencia de procesos. En primer lugar, la porción de manto que reviste la parte anterior del cuerpo en el extremo abierto de la concha secreta nuevo material de la misma. Así, el extremo abierto se hace a la vez más largo y más ancho, ampliando de ese modo la cámara corporal que contiene las partes blandas en crecimiento. Cuando el extremo abierto de la concha se ha desarrollado lo suficiente, el animal se desplaza hacia la parte anterior de la concha, dejando tras de sí un volumen de líquido cameral. El manto bloquea toda comunicación entre el animal y los tabiques de la cámara corporal. Una delgada capa de una sustancia viscosa puede reforzar ese aislamiento.

A continuación, el manto posterior secreta un nuevo septo calcáreo, empezando por los bordes de la concha y avanzando hacia dentro, hacia el sifúnculo, ubicado en posición central. Al mismo tiempo, el sifúnculo segrega su propio recubrimiento calcáreo. La geometría celular del sifúnculo de la cámara recién formada no es todavía del tipo que proponemos, apropiada para el vaciamiento de una cámara mediante ósmosis local. Sin embargo, en ese estadio el septo que se acaba de formar no posee todavía plena robustez, y el líquido de la cámara recién formada actúa de puntal contra la presión del agua de mar circundante, que se transmite al septo a través del cuerpo del nautilo. Si se extrajera el líquido antes de que el septo hubiera adquirido su consistencia definitiva, la nueva cámara se derrumbaría. Además, mientras aquél careciera de su vaina calcárea, la presión sanguínea



PARA VACIAR UNA CAMARA es preciso que las moléculas de agua atraviesen las capas externas porosas del sifúnculo y luego las células del epitelio sifuncular. La membrana de cada una de estas células epiteliales forma una red de finos repliegues. Estos repliegues comunican, a su vez, con un espacio extracelular mayor que puede servir como canal de drenaje, permitiendo que el agua penetre en los vasos sanguíneos del centro del sifúnculo. Posteriormente, los riñones del animal extraen el agua de la sangre y la excretan a la cavidad del manto. El vacío dejado por el agua se compensa en parte con la entrada de gas. El rectángulo de color muestra la parte de una célula epitelial que aparece en la ilustración inferior.

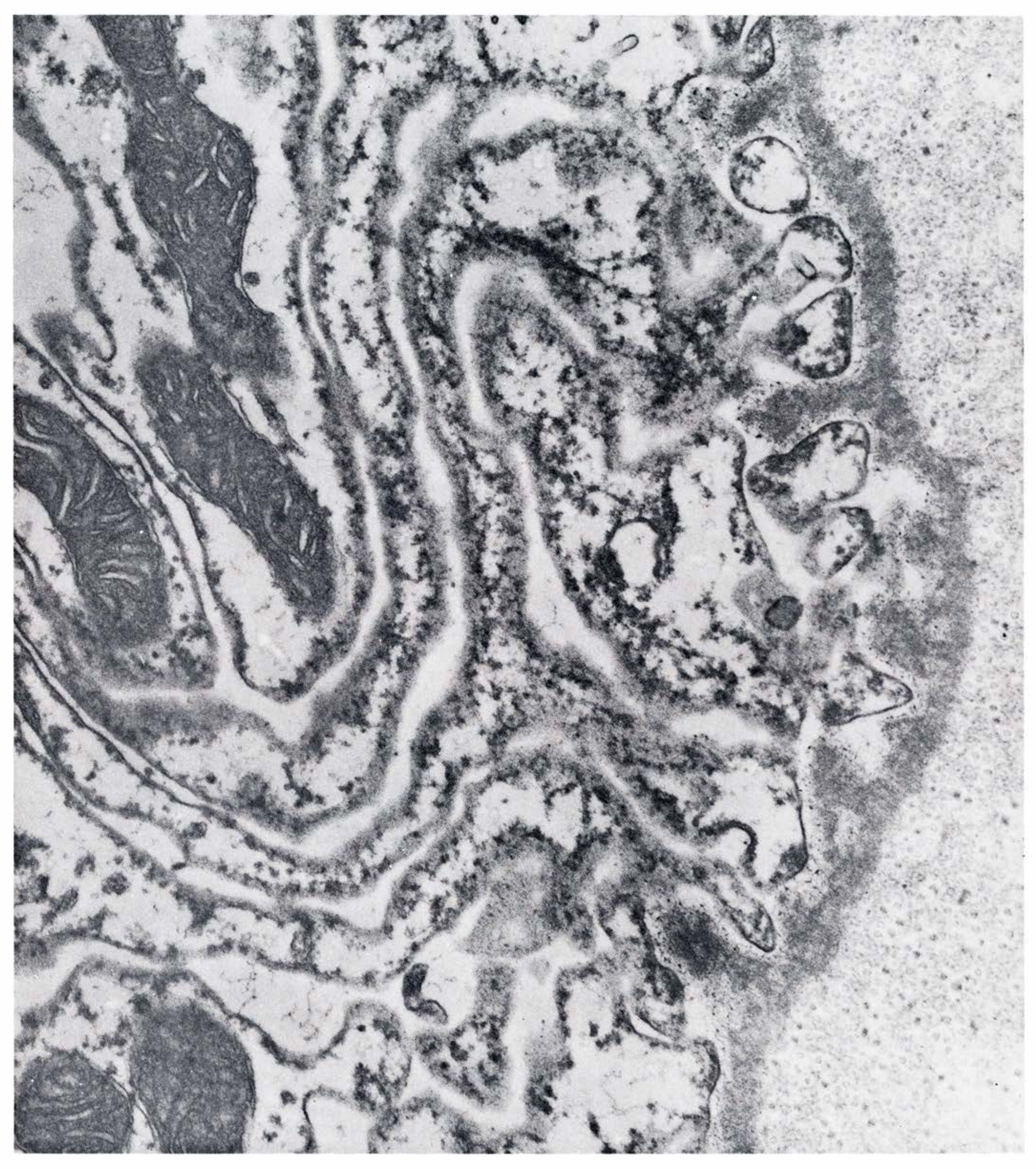


DETALLES DEL MECANISMO mediante el cual los autores proponen que se vacía la cámara aparecen en este dibujo esquemático de la red de finos repliegues de la membrana de una célula del epitelio sifuncular. El vaciado da comienzo cuando sustancias disueltas en el líquido cameral (*puntos de color*) se difunden en el interior del sifúnculo. Unos enzimas localizados sobre la membrana de las células sifunculares introducen el soluto hacia los repliegues. La energía para el bombeo viene suministrada por las mitocondrias, orgánulos intracelulares que proporcionan energía a las células. Cuando la concentración del soluto en la red de repliegues es lo bastante alta, el agua penetra por ósmosis (*flechas*) pese a la presión hidrostática de las profundidades marinas, que empuja el agua a salir del sifúnculo.

provocaría la explosión del sifúnculo si faltara el líquido cameral.

Terminado el septo, el sifúnculo presenta ya la estructura que consideramos idónea para la ósmosis local; la cámara empieza, entonces, a vaciarse. Cuando la cámara está llena contiene hasta 30 mililitros de líquido, mientras que la cámara contigua más reciente, y a veces las dos anteriores, cuentan con unos mililitros. Las cámaras más viejas están vacías. En aquellas cámaras donde el nivel del líquido ha caído por debajo del sifúnculo, puede extraerse también el líquido cameral. Llega al sifúnculo por capilaridad a lo largo de la superficie interior de la concha.

La eliminación de líquido de la cámara más reciente y de las adyacentes aumenta la fuerza de empuje del nautilo hasta un valor que contrarresta la flotabilidad perdida por la adición de peso a la concha y a los tejidos. Esta relación se hace patente al representar la longitud de la cámara corporal en función del volumen de líquido de la cámara inmediatamente posterior. Formada la cámara corporal, ésta es, en un comienzo, relativamente corta y, máximo, el volumen del líquido cameral.



MICROGRAFIA ELECTRONICA de una célula sifuncular, que confirma que las mitocondrias se congregan cerca de los finos repliegues. Las mitocondrias son los objetos más oscuros; en la parte izquierda de la imagen aparecen completas o en parte cinco de ellas. El canal extracelular de drenaje, más grande, mediante el cual los repliegues se comunican, corre de arriba abajo en el extremo derecho. El aumento es de 40.000 diámetros.

Conforme se alarga la cámara corporal, el volumen del líquido cameral disminuye. El ciclo se inicia cada vez que se forma una nueva cámara.

Pero el nautilo no crece indefinidamente. Como muchos animales, alcanza un determinado tamaño final, que coincide con la consecución de la madurez sexual. Su acercamiento a la madurez se manifiesta por una reducción en el espacio existente entre los dos o tres septos finales. Ese acortamiento varía bastante de un individuo a otro, y puede reflejar cierta capacidad del nautilo para efectuar ajustes finales de su flotabilidad. En cualquier caso, el encabalgamiento de los septos se aprecia no sólo en el nautilo, sino también en los retos fósiles de muchos cefalópodos tabicados. De ahí se deduce que muchos, si no todos los cefalópodos tabicados, tenían un límite fijo de desarrollo corporal y no crecían indefinidamente.

Concluido el último septo, el nautilo entra en los procesos finales que actúan en el ajuste de su flotabilidad. Vuelve a agrandarse el extremo abierto de la concha, espacio que ocupará el cuerpo del animal. Al mismo tiempo, los últimos restos del líquido cameral se eliminan de la cámara más reciente. El crecimiento ha terminado. El animal, con su concha incluida, mide varios centímetros de diámetro. La concha posee 30 o más compartimientos. El nautilo debe tener ahora por lo menos tres años de edad. El crecimiento de la concha y la secreción de los septos dentro de ella ha conducido a la consecución de la flotabilidad neutra, que perdurará en tanto el sifúnculo continúe achicando agua de la concha frente a la presión hidrostática del océano. Se desconoce cuánto tiempo vive el animal.

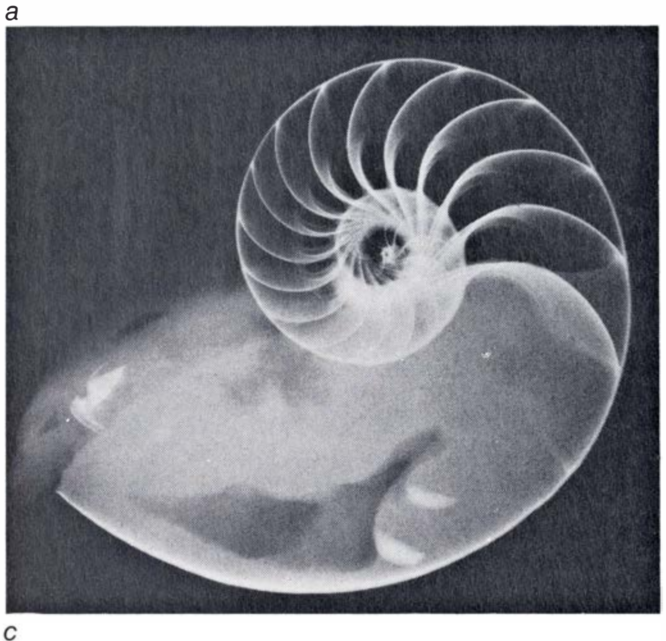
No sabemos todavía si existe algún mecanismo de retroacción biológica que mantenga sintonizado el crecimiento de la concha y el vaciado de las cámaras. De existir, observaremos, por ejemplo, que la formación de una nueva cámara se inhibe cuando evitamos que la cámara anterior se vacíe. Ya hemos señalado, que cuando el nivel del líquido de la cámara más reciente cae por debajo del nivel del sifúnculo, comienza la secreción de un nuevo septo.

Además, hemos demostrado que el nautilo puede compensar cambios artificiales en su flotabilidad. En nuestros primeros experimentos nos enfrentábamos con el problema de tasas de vaciado tan bajas que apenas podían medirse. Por ello ofrecíamos a nuestro animales un “incentivo” para vaciar la

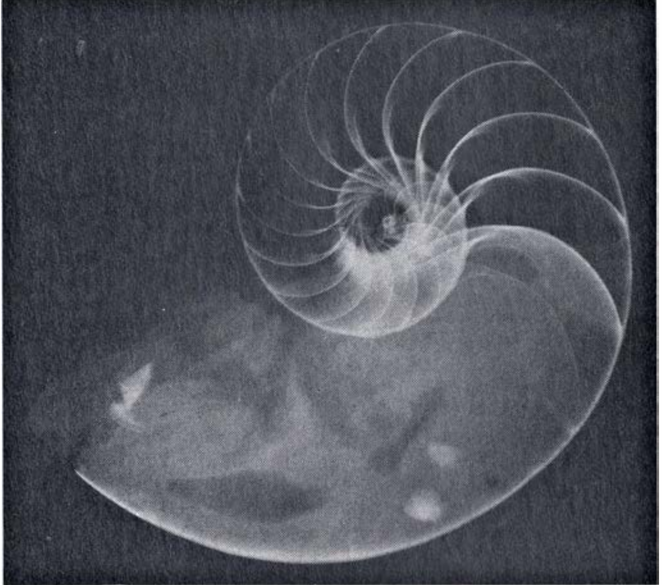
cámara experimental (aquélla en la que efectuábamos las mediciones): añadíamos líquido a una segunda cámara, vacía, con lo que aumentábamos el peso del animal en el agua. En todos los animales sometidos a un sobrepeso por haberles inundado una cámara, observamos un vaciado mucho más rápido de la cámara experimental. Cuando, más tarde, provocamos en los mismos una fuerza de empuje desmesurada por achique de la cámara inundada, descubrimos una reducción de la velocidad de evacuación de líquido. El vaciado más rápido observado en el laboratorio fue de menos de un mililitro por día, valor que corresponde a la estima de que el animal precisa por lo menos un mes para vaciar una cámara.

Puesto que en todas nuestras observaciones del nautilo en aguas someras no hemos observado nunca que un animal añadiera agua a una cámara antes vacía, concluimos que sólo dispone de una opción: aumentar o disminuir la tasa de vaciado de la cámara (y quizás el ritmo de formación de la concha). ¿Por qué razón, pues, mantiene su sifúnculo vivo en todas las cámaras, incluso en la primera? Sacaría más partido de sus posibilidades si, en cambio, hubiera desarrollado un mecanismo para cerrar herméticamente las cámaras vacías. Quizá la porción de sifúnculo que se extiende hasta las primeras cámaras sea el vestigio de un órgano capaz de inundar y vaciar las cámaras en los cefalópodos tabicados ancestrales. Estos animales, hoy ya extintos, podían haber sido así mucho más móviles que aquellos a los que han dado lugar por evolución.

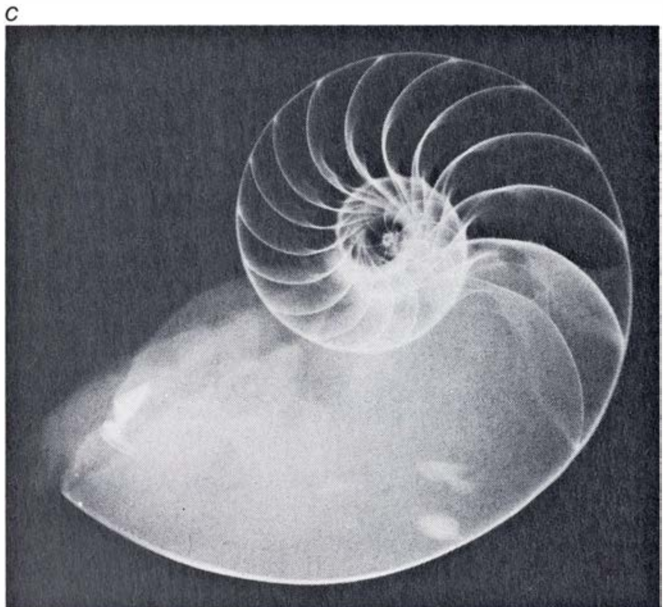
a



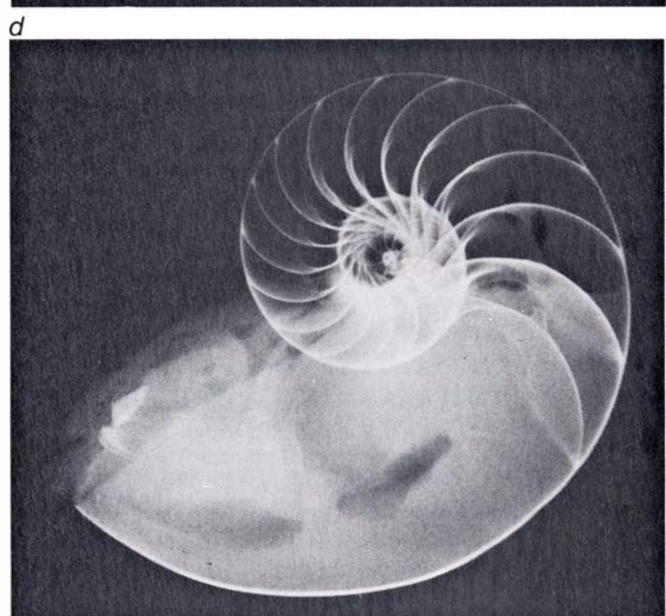
b



c



d



SECRECION DE UN SEPTO y vaciado de las cámaras en la concha. Ambos procesos pueden seguirse en esta secuencia de cuatro radiografías de un nautilo vivo. En *a* el nivel del líquido de la cámara más reciente ha caído por debajo de la base del sifúnculo. En la cámara que se encuentra por encima queda una cantidad mucho menor de líquido. Todas las demás cámaras están vacías. En *b* (realizada tres semanas después de *a*) existe un nuevo septo vagamente insinuado. La cámara que define está llena de líquido cameral. Mientras tanto, el nivel del líquido en lo que había sido la cámara más reciente ha continuado bajando. Aunque el líquido de esta cámara no está ya en contacto con el sifúnculo directamente, puede alcanzar a éste al desplazarse por capilaridad a lo largo de la superficie interna de la concha. En *c* (realizada dos semanas después de *b*) el septo más reciente se encuentra casi terminado. En *d* (realizada dos semanas después de *c* y siete semanas después de *a*) el septo más reciente está finalizado. Las dos regiones oscuras de la zona superior izquierda de la cámara recién formada son burbujas de gas. Su presencia indica que la cámara más reciente se está vaciando. A lo que parece, el nuevo septo ha adquirido fuerza suficiente para soportar la presión oceánica sin la abrazadera hidrostática que proporciona el líquido cameral. En cada una de las imágenes el nautilo vivo ocupa la cámara corporal, en la parte inferior de la concha. Los dos puntos blanquecinos a la derecha del cuerpo en *a* son piedras nefríticas, que han desaparecido en *d*. Las manchas blancas a la izquierda de cada imagen son las mandíbulas del cefalópodo.

Juegos matemáticos

De contar los votos a lograr que cada voto cuente: sobre matemática electoral

Martin Gardner

Tras las elecciones presidenciales de los Estados Unidos del mes pasado puede ser oportuno echar una ojeada sobre los aspectos matemáticos de las votaciones, aunque no fuera más que por las diversas paradojas y anomalías de origen estrictamente matemático que han podido tener importancia en esta confrontación ante las urnas. Veremos aquí en qué forma estas contradicciones afectan a la elección por mayoría y al trámite del colegio electoral, y veremos también un sistema de votación relativamente nuevo, conocido por voto de aceptación. Este procedimiento, que cada día interesa más a los especialistas en ciencias políticas, consigue eludir muchas incoherencias de carácter lógico que presentan otros sistemas de votación.

El análisis que este mes publicamos sobre matemática electoral no ha sido escrito por mí, sino por Lynn Arthur Steen, profesor de matemáticas del St. Olaf College, de Northfield, Minnesota. Steen, que es redactor jefe de *Mathematics Magazine,* suele dedicar su pluma a temas de matemáticas, y ha sido galardonado por dos veces con el premio Lester R. Ford, de la Mathematical Association of America, por la perfección de su estilo. Ha compilado diversidad de libros, entre ellos *Mathematics Today: Twelve Informal Essays* (Springer-Verlag, 1978) y (en colaboración con Matthew P. Gaffney) *Annotated Bibliography of Expository Writing in the Mathematical Sciences* (editado por la Mathematical Association of America, 1976).

Todo cuanto sigue, hasta los comentarios finales sobre secciones de meses anteriores, que son míos, lo ha redactado Lynn Arthur Steen, quien ha titulado así su trabajo sobre anomalías y paradojas electorales: "Election Mathematics: Do All Those Numbers Mean What They Say?" (Matemáticas electorales. ¿Quieren decir los números lo que parecen?)

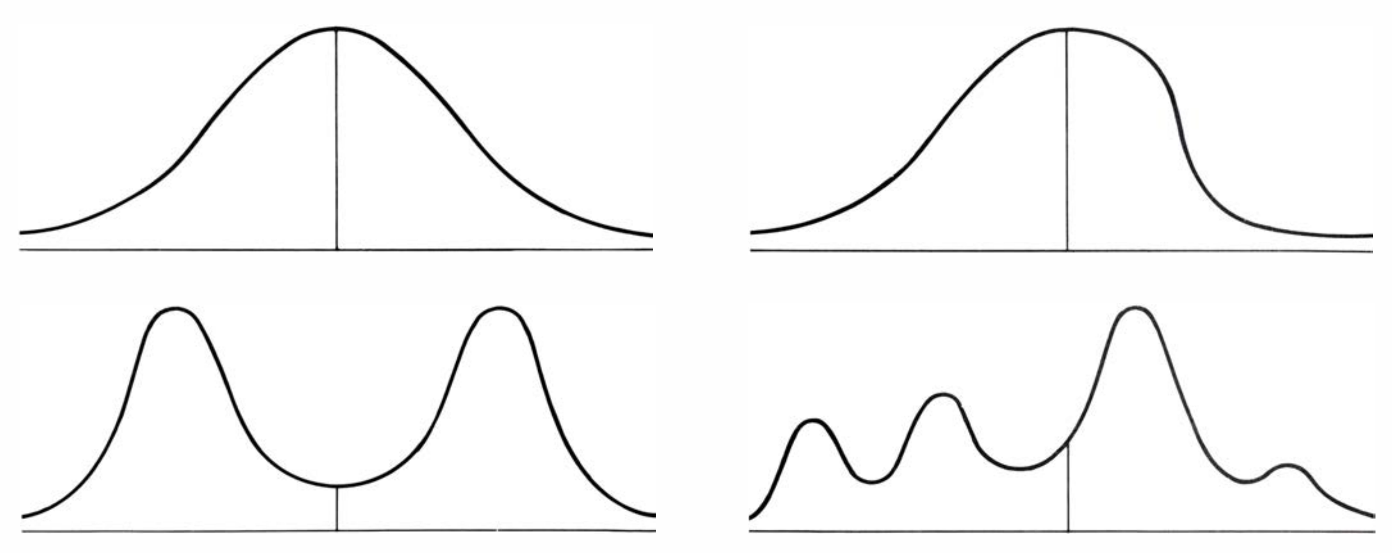
A lo largo de todo el año pasado, la perspectiva de que la carrera electoral hacia la Casa Blanca tuviera tres competidores ha hecho recaer la atención pública sobre la importancia de las estrategias de voto y sobre las peculiaridades del colegio electoral. Y aunque las complicaciones impuestas por el colegio electoral se refieren únicamente a elecciones presidenciales, no sucede así en otros tipos de incertidumbres provocadas por candidaturas múltiples a cargos públicos. Cuando los votantes han de optar entre varias candidaturas, la tarea de formarse la decisión final puede ser exasperantemente difícil. La fuente tanto de estas dificultades como de sus posibles soluciones ha de buscarse en la teoría matemática de elecciones, tan mal conocida.

El contrato social de toda democracia se funda básica y obviamente en una sencilla idea matemática, a saber, la noción de mayoría. Dejando de lado el muy improbable caso de un empate, en toda votación dicotómica uno u otro bando habrá de recibir más de la mitad de votos. En cambio, cuando hay tres o más candidaturas de fuerzas sensiblemente iguales será improbable que sufragios de este tipo consigan producir decisiones mayoritarias. Si tantas personas creen que el sistema bipartidista es esencial para la estabilidad de la democracia en los Estados Unidos es sobre todo por este motivo, a pesar de que tal sistema no está ni decretado ni especialmente reconocido en la Constitución.

Mas digan lo que dijeren la teoría matemática y el idealismo político, es frecuente que el público tenga que elegir entre tres o más candidaturas importantes. El mismo problema que se presentó este año en la pugna "Carter contra Reagan contra Anderson" se ha presentado ya en otras ocasiones. Es difícil resolver pulcramente estas confrontaciones de candidaturas múltiples cuando no se producen mayorías bien definidas; empero, es fácil que se produzcan en toda elección libre. Más aún, sencillos razonamientos hacen ver que incluso en un duelo a dos apenas hay en la práctica posiciones políticas que hagan a los candidatos invulnerables frente a los ataques de terceros o cuartos aspirantes.

Supongamos que, en una elección entre dos candidatos, las preferencias del electorado sobre cada uno de los puntos a debate de sus programas vinieran representadas mediante un valor numérico sobre una escala unidimensional. Resulta entonces que, independientemente de la distribución de actitudes entre el electorado, la posición óptima para cada candidato se encuentra en el valor llamado "mediana", punto que divide al electorado en dos campos de igual tamaño. Nada importa que la opinión pública siga su distribución normal (la línea que representaría el número de votos favorables en función de la posición política del candidato tendría una sola prominencia, y sería simétrica respecto a ésta), que esté fragmentada bimodalmente (la gráfica tendría dos prominencias sensiblemente iguales), que esté fuertemente sesgada hacia un lado, o que esté distribuida de forma irregular. En la ilustración adjunta puede verse un ejemplo de cada una de estas distribuciones, así como la situación de la mediana [*véase la ilustración superior de la página contigua*].

Tomemos unos comicios con dos candidatos, donde uno de ellos adopte una posición situada ligeramente a la izquierda de la mediana y el otro comience sensiblemente en el centro del ala derecha de la población. Este sería típicamente el caso de un candidato centrista *C* enfrentado a un aspirante perteneciente a la derecha moderada *D*. En esta hipótesis, sería razonable suponer que (en lo tocante a este punto concreto del programa) los votantes con preferencias situadas a la izquierda del candidato centrista *C* favorecerían



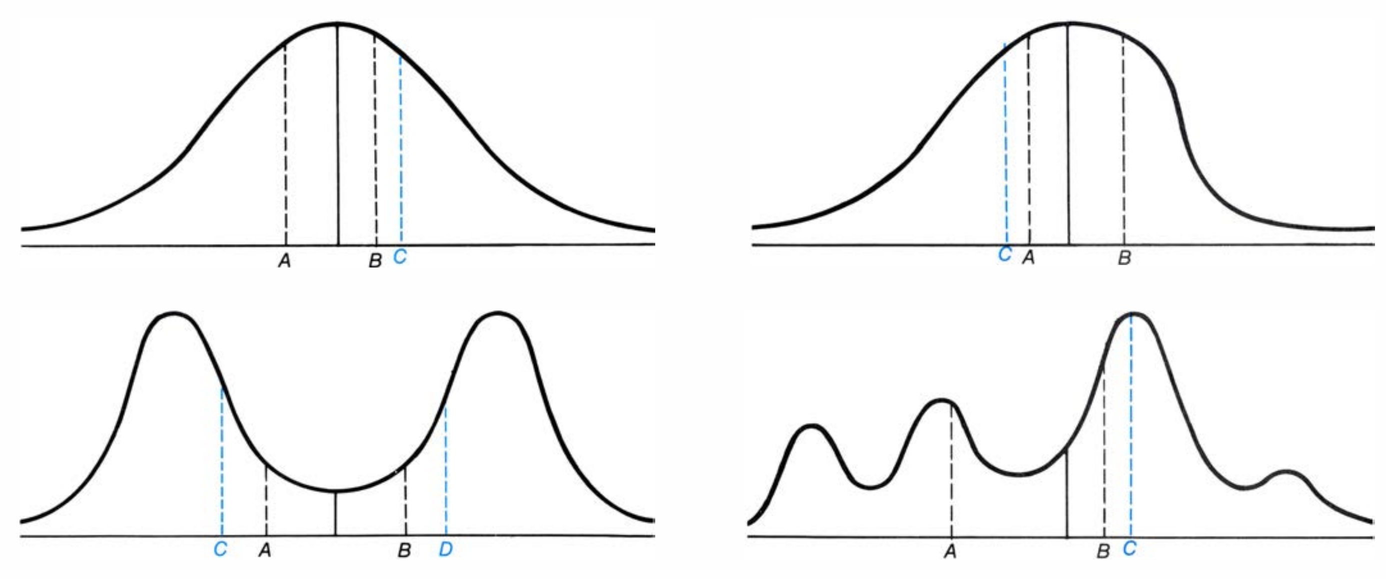
Cuatro distribuciones posibles de la opinión pública

a *C,* que los electores cuyas preferencias se encuentran a la derecha de *D* votarán a *D,* y que los votantes de tendencias intermedias, situados entre ambos candidatos, se dividirán en dos grupos sensiblemente iguales uno para cada uno. Así las cosas, en un sondeo pre-electoral el candidato centrista recibiría la mayoría de los votos.

La única forma de que el candidato *D* mejore su porción de electorado (tan sólo en lo relativo a este punto concreto del debate) es desplazar su posición hacia el centro de la distribución, para asegurarse de que a su derecha quedará mayor número de votantes. Desplazarse hacia el centro, o sea, hacia la izquierda, siempre será beneficioso para el candidato de la derecha. Análogamente, un candidato de tendencia izquierdista puede mejorar su puesto frente a los electores deslizándose hacia el centro, que, en su caso, es ocupar posiciones más a la derecha. Unicamente cuando se ocupe la posición mediana será imposible mejorar mediante nuevos desplazamientos.

Como es obvio, este análisis no contiene nada verdaderamente novedoso. Forma parte de la experiencia adquirida ya en muchas elecciones presidenciales. En general, los candidatos representantes de la derecha y la izquierda tienden a ocupar inicialmente posiciones bien caracterizadas, para luego ir progresivamente deslizándose hacia el centro, intentando así captar mayor número de votos. Empero, precisamente por ser la posición mediana tan atractiva para ambos candidatos, resulta muy vulnerable pudiendo ser atacada desde los flancos por terceras o cuartas candidaturas. En cualquier pugna entre dos candidatos próximos al centro, un tercero que entre en lid, bien por la derecha, bien por la izquierda, puede lograr la mayoría. Más todavía; en contiendas entre dos candidaturas no hay posición alguna donde al menos una de ellas no pueda ser derrotada por una tercera, y ello para prácticamente todas las distribuciones de opinión en el electorado. Como vemos en la ilustración contigua [*ilustración inferior de esta misma página*], a lo largo del continuo unidimensional existe siempre un lugar donde un nuevo candidato allí situado desplazaría a uno o más rivales cercanos.

Muy raras veces tiene un solo punto del debate papel decisivo en las elecciones. Por tanto, los análisis electorales basados en ellos no sirven de mucho, a menos que sepamos combinarlos para diseñar un programa electoral capaz de traer la victoria al candidato. Pero

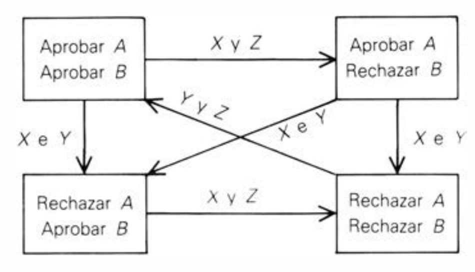


*Dos cantidatos próximos al centro (***A** *y* **B***) pueden ser vencidos por un tercero (***C***) y a veces un cuarto (***D***) que entren en liza*

Alternativas

I. Aprobar las propuestas *A* y *B*.
II. Aprobar la propuesta *A* y rechazar la *B*.
III. Rechazar la propuesta *A* y aprobar la propuesta *B*.
IV. Rechazar las dos propuestas *A* y *B*.

Sector	Opinión	Votos	Orden de preferencia
X	Fuertemente favorable a *A*, ligeramente favorable a *B*	150	I, II, III, IV
Y	Francamente contraria a *B*, ligeramente favorable a *A*	150	II, IV, I, III
Z	Decididamente contraria a *A*, ligeramente favorable a *B*	200	III, IV, I, II



Tres facciones, al votar dos puntos programáticos, pueden crear mayorías cíclicas

atinar en semejante programa es asunto delicado, por existir la posibilidad de que salgan derrotados programas que tomados punto a punto, aisladamente, resultarían vencedores. El reverso de esta moneda nos muestra ser posible construir programas capaces de lograr apoyo mayoritario, compuestos por elementos aisladamente minoritarios. Y ello, porque puede lograrse la mayoría mediante coalición de minorías.

Para ver cómo puede presentarse esta paradoja examinemos el caso más sencillo posible: una votación que deba decidir entre dos cuestiones independientes y dicotómicas, representadas por las propuestas *A* y *B*. En este caso, los votantes tienen cuatro opciones:

I. Aprobar las propuestas *A* y *B*.
II. Aprobar *A*, pero rechazar *B*.
III. Rechazar *A*, y en cambio apoyar *B*.
IV. Rechazar *A* y rechazar *B*.

Los electores partidarios de *A* y de *B* colocarían la opción I en primer lugar, la opción IV en cuarto lugar, y la opción II en segundo o tercer lugar, según juzgasen más importante la propuesta *A* o la propuesta *B*. Los votantes favorables a *A*, pero contrarios a *B* ordenarían las opciones en la forma II, I, IV y III (o tal vez II, IV, I y III). En general, cada elector se inclinaría por una de las 4 × 3 × 2 × 1 (o sea, 24) posibles permutaciones de las cuatro opciones disponibles. (Ello no quiere decir que las permutaciones sean todas igualmente verosímiles; sería difícil imaginar circunstancias en que mucha gente adoptase el orden de preferencia I, IV, II, III.)

Por sencillez, supongamos ahora que 500 electores (pongamos por caso, los compromisarios de la convención de un partido) estén divididos en tres sectores, como sigue: el grupo *X*, que dispone de 150 votos, sitúa las cuatro opciones en el orden I, II, III y IV; el grupo *Y*, con 150 votos, las prefiere en el orden II, IV, I, III; y el grupo *Z*, con 200 votos, en el orden III, IV, I, II. En este caso, los grupos *X* e *Y*, que reúnen 300 votos, serían favorables a la propuesta *A*, mientras que *X* y *Z*, con un total de 350, apoyarían la aprobación de la propuesta *B*. Mas siendo distintos los votantes que en cada caso componen estas mayorías, ¡la candidatura formada por la propuesta "aprobación conjunta de *A* y de *B*" sería derrotada por el bloque de 350 votos reunidos por los sectores *Y* y *Z*!

Tan sorprendente fenómeno es caso particular de la conocida paradoja de las mayorías cíclicas: Si tres votantes prefieren, respectivamente, *A* a *B* a *C*, *B* a *C* a *A*, y *C* a *A* a *B*, resultará que cualquier candidato puede ser derrotado por algún otro candidato por dos votos contra uno en toda pugna entre dos. Cuando en unas elecciones los puntos a debate suscitan mayorías cíclicas, ningún sistema de posiciones que puedan adoptar un par de candidatos puede hacerlos invulnerables al asalto de nuevas coaliciones minoritarias, siendo este otro de los factores que pueden decidir a terceras o cuartas partes a presentar sus candidaturas.

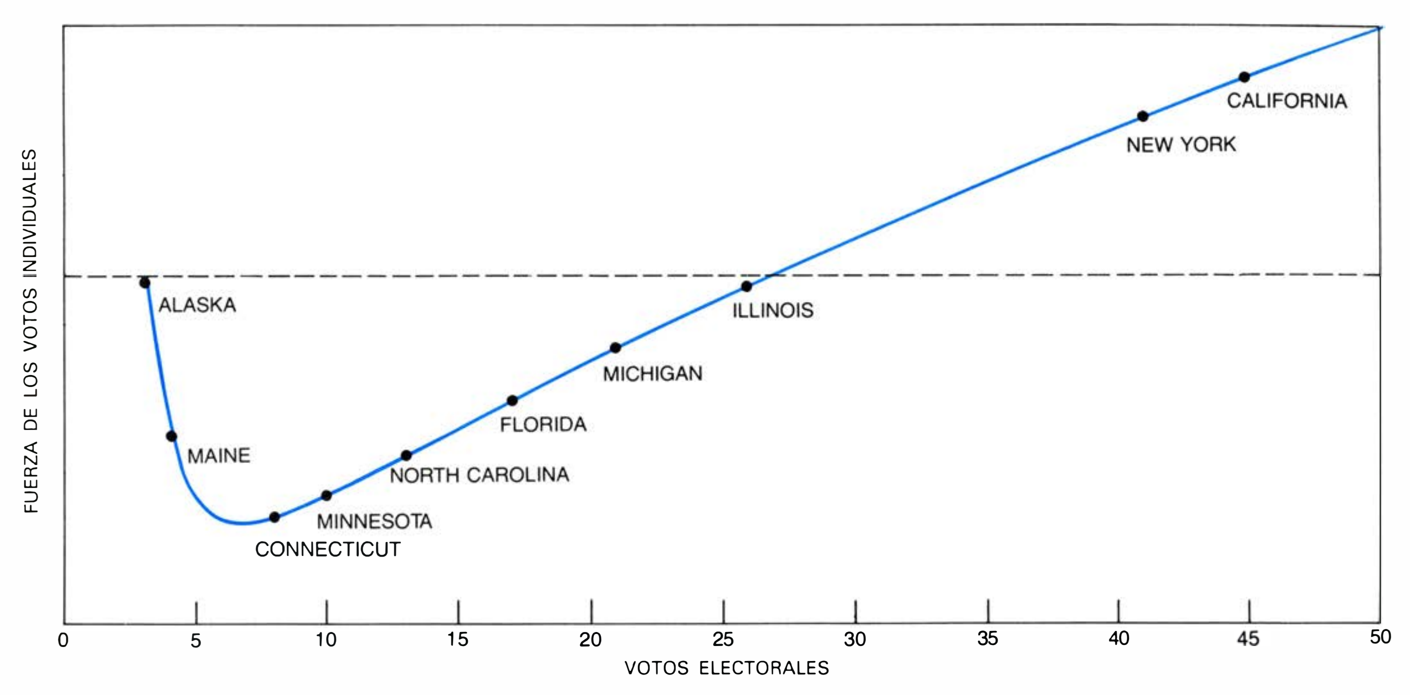
El diagrama adjunto [*véase la ilustración superior de la izquierda*] muestra cómo las cuatro opciones con las que deberán confeccionarse los programas electorales del partido van creando diversas mayorías cíclicas, y ello explica por qué una candidatura formada por propuestas individualmente mayoritarias puede quedar globalmente respaldada sólo por una minoría. Las flechas que conectan las distintas candidaturas denotan las relaciones de dominancia electoral: la candidatura situada en la punta de flecha resultaría derrotada por la situada en el origen de ésta en caso de quedar enfrentadas sólo estas dos. En cada caso, junto a cada flecha están anotados los sectores victoriosos. Como esta distribución pone de manifiesto, toda posible candidatura puede ser derrotada por alguna otra, y así, si las facciones del congreso de un partido estuvieran divididas de forma semejante a las de nuestro ejemplo, la convención podría empantanarse en una serie interminable de mociones, que irían cada una derrotando a su precedente.

También el fenómeno de las mayorías cíclicas es responsable de la más famosa de las paradojas electorales. En 1951, Kenneth J. Arrow demostró que ciertas características comúnmente consideradas, muy deseables en los sistemas de votación, son incoherentes

Orden de preferencia	Total de votos	Votos de aprobación: Primer lugar	Votos de aprobación: Primer y segundos lugares
A, B, C	30	20	10
A, C, B	5	5	0
B, A, C	20	10	10
B, C, A	5	5	5
C, A, B	10	5	5
C, B, A	30	20	10
Total	100	65	35

Elección por mayoría	Elección a dos vueltas	Elección por puntuación	Elección por aprobación
A 35	*A* 35 + 20 = 55	*A* 35 (3) + 30 (2) + 35 = 200	*A* 25 + 10 + 15 = 50
B 25	*C* 40 + 5 = 45	*B* 25 (3) + 60 (2) + 15 = 210	*B* 15 + 10 + 20 = 45
C 40		*C* 40 (3) + 10 (2) + 50 = 190	*C* 25 + 15 + 0 = 40

Comparación de métodos electorales para el caso de tres candidatos



Representación gráfica de la fuerza del voto individual en varios estados

desde el punto de vista lógico. No surgen problemas cuando sólo se presentan dos candidatos; mas cuando en una sola votación es preciso optar por uno de tres o más candidatos, el caos impera.

Para determinar el vencedor de unas elecciones hay, además del sistema de mayoría, otros varios; muchos de ellos propuestos por estudiosos del siglo XVIII, preocupados por lograr instrumentos capaces de plasmar y materializar los principios democráticos de la Revolución Francesa. Si bien la complejidad de algunas de estas propuestas las haría impracticables, otras se encuentran hoy en pleno uso; entre ellas, el método de asignación de puntos que vayan traduciendo el distinto grado de preferencia del electorado por los candidatos (siendo vencedor el candidato que llegue a reunir mayor número de puntos), así como diversos métodos para celebrar elecciones a dos vueltas, una de ellas eliminatoria. Empero, como ha demostrado Arrow, ninguno de estos sistemas –en realidad, ningún otro sistema distinto de una dictadura racional y benévola– es capaz de satisfacer reglas tan de sentido común como: Si *A* es preferido a *B*, y *B* es preferido a *C*, entonces *A* ha de ser preferido a *C*. La aparición de mayorías cíclicas se encarga de transformar todo plan de votación en un misterio impredictible.

Otro importante problema que se presenta en elecciones con tres o más opciones es que, en muchos casos, al votar el elector por su candidato preferido no hace sino aumentar la probabilidad de que resulte electo el candidato para él menos deseable. (Este dilema ha podido presentarse a causa de la candidatura de Anderson. Muchos votantes que preferirían votar por Anderson mejor que por Carter, y por éste mejor que por Reagan, podrían temer que casi todos los votos de Anderson hayan sido detraídos a Carter.) Esta anomalía suele llevar a los electores reflexivos al llamado voto útil.

De generalizarse el voto insincero, podría llegarse a un estado de grave confusión, donde nadie votaría a su preferido, y donde la voluntad del pueblo quedaría desvirtuada por completo. Un partidario de Anderson, para quien Carter fuese un mal menor, pudiera haber votado por Carter en lugar de por Anderson, para impedir la elección de Reagan. Si hubiera un número importante de partidarios de Anderson que procediesen de esta forma, es evidente que algunos partidarios de Reagan pudieran haber votado a Anderson en las elecciones primarias pensando en evitar la reelección de Carter. Pero el proceso de segundas intenciones, pensando en la estrategia de voto de otros sectores del censo, puede rápidamente conducirnos a una absurda jerarquía de insinceridades, y los votos dejarían de reflejar las verdaderas preferencias del electorado. Semejante proceso, que preciso es decirlo, más forma parte de la teoría de juegos que de la teoría clásica de elecciones, raramente puede darle al ganador un mandato legítimo.

El teorema de Arrow muestra que no existe un sistema de votación "perfecto" cuando concurren varios candidatos. El procedimiento conocido por "voto de aprobación" o "voto de consentimiento" consigue empero reflejar una voluntad popular sin inducir a nadie a votar contra su preferencia. En el voto de aprobación cada elector marca sobre la papeleta todos los candidatos que merecen su aprobación; el candidato que reciba el mayor número de votos es el vencedor.

Con este sistema, al votante nunca puede resultarle ventajoso privar de su apoyo al candidato de su preferencia, votando en cambio a otro candidato menos deseable. Más aún, si las probabilidades de vencer fuesen razonablemente iguales para la mayoría de los candidatos, un elector racional debería dar su voto a todos los candidatos que en su opinión se encuentren por encima del promedio de competidores. Votar por más candidatos concedería innecesario apoyo a individuos que el votante no estaría dispuesto a respaldar, mientras que votando por menos (por ejemplo, solo a su candidato favorito) se le retira apoyo a candidatos que pudieran ser aceptables, aumentando al mismo tiempo el riesgo de que venza un candidato inaceptable.

Stephen J. Brams, profesor de ciencias políticas de la Universidad de Nueva York, ha descrito el voto de aprobación con la feliz frase "Un hombre, *n* votos". Descripción verdaderamente lúcida, porque el voto de consentimiento no es sino un método

para permitirle a cada elector votar tantas veces como desee, una por cada candidato aceptable. El recuento de votos es sencillo en este sistema, y no es preciso recurrir a votaciones eliminatorias o a segundas vueltas. Tanto por razones de orden práctico como de índole teórico, el voto de aprobación es un excelente compromiso entre la elección a simple vuelta, que puede obligar al elector a emitir votos insinceros, y la ordenación completa por grado de preferencia, cuyo escrutinio lo hace impracticable por su complejidad.

La ilustración inferior de la página 124 expone comparativamente los comportamientos de las elecciones por mayoría, a dos vueltas, por puntuación, y por aprobación en una lid electoral enteramente hipotética con tres candidaturas. El número de votos que respaldan a cada una de las seis posibles ordenaciones de los contendientes están reseñados en la columna "total de votos", y puesto que el bloque donde *C* encabeza la lista recibiría el máximo de votos, este candidato sería vencedor en una elección por mayoría simple. Si la elección fuese a dos vueltas, *B* quedaría eliminado, y *A* recibiría suficientes votos (al figurar como preferido en segundo lugar y recibir en la segunda vuelta los votos de quienes antes votaron por *B*) para vencer a *C* por 55 votos contra 45. En el más sencillo de los sistemas de votación por puntos al preferido en primer lugar se le asignan tres puntos, dos al preferido el segundo, y uno al tercero. A causa del gran número de electores (60) que optarían por *B* en segundo lugar, con este sistema electoral el candidato *B*, que por el método a dos vueltas quedaría eliminado en la primera, sería ahora el vencedor.

Los resultados del sistema de votación por consentimiento dependerían de que los electores sólo encontrasen aceptable a su candidato favorito, o de que admitieran también a otros. (Al haber únicamente tres candidatos, se supone en este ejemplo que nadie votaría por los tres a la vez. Tales votos, aunque válidos, serían desperdiciados, pues incrementarían los totales de cada uno de los candidatos en la misma cantidad.) En este caso, donde hay 65 electores que solamente dan su aprobación a su candidato favorito, *A* recibiría 50 votos aceptándolo, y ganaría la elección. Si algunos votantes decidiesen aceptar a dos de los tres aspirantes, es probable que ganase *B*, en razón, sobre todo, del gran número de votos que lo sitúan como segundo. En el sistema de aprobación puede bastar que un pequeño número de votantes varíe el número de candidatos considerados aceptables para cambiar el resultado de los comicios. Por tanto, llevar a la práctica este sistema electoral exigiría modificar la estrategia de la campaña electoral, y en lugar de proponerse convencer al electorado de que el candidato es óptimo, debería intentar hacerle ver que es aceptable.

Como es sabido, en los Estados Unidos la elección presidencial se rige por las reglas del colegio electoral, que difieren totalmente de lo expuesto. A lo largo de casi toda la historia norteamericana, el colegio electoral ha servido primordialmente para imponer a cada estado una regla unitaria, de forma que el vencedor del voto popular en cada estado reciba la totalidad del voto electoral del estado. De acuerdo

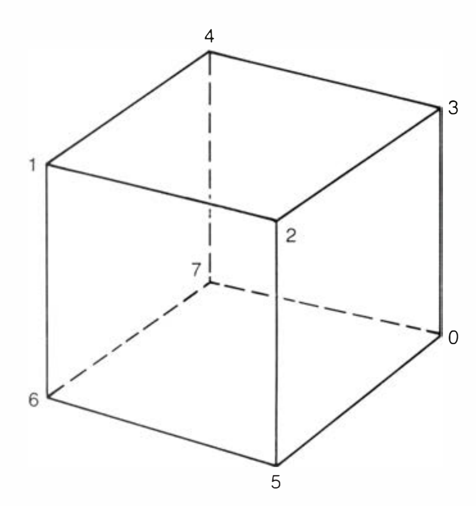
con la Constitución de los Estados Unidos, este sistema tiene otras importantes consecuencias, que pueden afectar al desenlace de elecciones de triple candidatura (en concreto, existen disposiciones que transfieren la responsabilidad de la decisión final del colegio electoral al Congreso). Aquí nos limitaremos a examinar las consecuencias de la regla unitaria.

Con respecto a la regla unitaria impuesta por el colegio electoral, la opinión más difundida es que favorece a los estados más pequeños, o mejor, a los menos poblados, debido a que el número de votos que el colegio electoral atribuye a cada estado es igual al número de representantes, más dos. En términos comparativos, estos dos votos extra correspondientes a los dos senadores de cada estado refuerzan el valor del voto individual de los electores de estados pequeños, y lo disminuyen en los estados grandes.

Paradójicamente, sin embargo, en las elecciones presidenciales la fuerza efectiva de cada estado es en realidad proporcional a la potencia de exponente 3/2 de la población del estado. Y, en consecuencia, la importancia de los votos individuales emitidos en estados grandes puede llegar a ser tres veces mayor que en los más pequeños. Tan sorprendente conclusión se deduce de la teoría elemental de probabilidad, y concuerda con las cantidades que los candidatos han invertido en elecciones recientes: los aspirantes destinan fracciones desproporcionadamente grandes de sus recursos a los estados más poblados, a expensas de los menos habitados.

La "regla de los 3/2" se funda en la hipótesis de que, por lo común, en sus campañas electorales por los distintos estados los candidatos en liza tratan de igualar o superar las de sus rivales. (Comparando la distribución de tiempo y dinero de campañas recientes se comprueba que esta presunción se ajusta a la realidad.) El razonamiento arranca de algo evidente: cada candidato se esfuerza por hacer máximo su voto electoral esperado, que resulta de sumar para los 50 estados el producto del voto electoral del estado por la probabilidad de que el candidato alcance mayoría relativa en el estado. Traduciendo esta relación a una fórmula matemática y tomando en consideración la tendencia de los candidatos a igualar, estado por estado, las campañas de sus rivales, puede demostrarse que el procedimiento óptimo para hacer máximo el voto electoral esperado consiste en repartir los recursos disponibles para la campaña en proporción aproximada a la potencia 3/2 del voto electoral del estado. Así resulta que aunque California tiene aproximadamente cuatro veces el voto electoral de Wisconsin (45 votos frente a 11), la regla de los 3/2 indica que los candidatos deberían dedicar $4^{3/2}$, es decir, 8 veces más recursos a California que a Wisconsin.

Otra forma de comprender por qué los estados grandes no resultan debilitados, sino, por el contrario, fortalecidos por la estrategia del colegio electoral consiste en examinar la probabilidad de que un voto individual puede ser decisivo para volcar los votos del estado en favor de uno u otro candidato. Tradicionalmente se ha tomado esta estimación como método para valorar la fuerza del elector individual. Lo que



Forma de rotular los vértices de un cubo, problema que se planteó el mes pasado

se requiere es una medida del número medio de votos necesarios para dar la vuelta al resultado de una elección en cada estado.

Los cálculos hacen ver que la capacidad decisoria de un individuo de un estado que disponga de ν votos electorales (que serán emitidos en bloque en el colegio electoral) es proporcional a $\sqrt{\nu}$. Como en el colegio electoral tal capacidad decisoria queda multiplicada por el número de votos electorales emitidos por el estado, el peso relativo de cada estado en la decisión final es aproximadamente proporcional a ν veces $\sqrt{\nu}$, o sea, a $\nu^{3/2}$.

A la hora de valorar comparativamente la fuerza del voto de los individuos de unos y otros estados es necesario ponderar por una parte la distorsión creada por la regla de los 3/2, que favorece a los estados grandes, y por otra, la bonificación de los dos senadores por estado, que favorece a los pequeños. El valor del voto de cada individuo, en lugar de ser el mismo para todos, es en realidad parte alícuota del peso electoral del estado en que habita, y, como muestra la ilustración de la página 125, la fuerza del voto individual en los distintos estados es francamente desigual. (La línea de trazos del gráfico señala la distribución que sería hipotéticamente justa.)

Las elecciones seguirán siendo siempre asuntos más pasionales que lógicos, y basándose más en la convicción que en la razón. Mas como estos ejemplos hacen ver, la matemática electoral puede tener sutiles e insospechadas consecuencias. Y como en otros muchos dominios de la experiencia humana, estructuras matemáticas sencillas, encubiertas en forma de paradojas y anomalías, pueden echar por tierra las ilusiones más ingenuas.

Con lo anterior concluye el artículo de Steen. Las soluciones de los problemas del mes pasado son como sigue:

Los cuatro precios que Iva pagó por sus baratijas fueron un dólar, un dólar y medio, dos dólares y dos dólares y veinticinco centavos. Tanto la suma como el producto de estos números es 6,75 dólares. De no haberse especificado que Iva pagó un dólar por los pendientes, hubiera existido una segunda solución: un dólar y veinte centavos, un dólar y veinticinco centavos, un dólar y ochenta centavos y 2 dólares y medio. Este problema, propuesto por Kenneth M. Wilke, es variante de otro, preparado por David A. Grossman, que figura como problema n.º 65 en *Ingenious Mathematical Problems and Methods,* de L.A. Graham (Dover Publications, 1959). En la versión de Grossman tanto la suma como el producto son 7,11; y el único conjunto de precios es un dólar y veinte centavos, un dólar y veinticinco centavos, un dólar y medio y 3 dólares y 16 centavos.

Un segundo método para descomponer un cubo en tres sólidos congruentes consiste, sencillamente, en cortarlo en tres rebanadas paralelas iguales. No conozco ya un tercer procedimiento para cortar el cubo en tres piezas idénticas. La ilustración de esta página muestra la única forma de rotular los vértices de un cubo con los números de 0 a 7 (despreciando las variantes deducidas por giros o simetrías) de modo que la suma del par de números situados en los extremos de cada arista sea número primo. No cuesta mucho demostrar que no existe ninguna otra solución con ocho dígitos consecutivos mayores que 0.

La paradoja de las razones de superficie a volumen en cubos y esferas está en el primer problema de *Science Brain-Twisters, Paradoxes and Fallacies* por Christopher P. Jargocki (Charles Scribner's Sons, 1976). La falacia del razonamiento se debe a que el símbolo *d* ha recibido dos significados enteramente distintos. La forma correcta de comparar las razones de superficie a volumen de la esfera y el cubo es tomar dos cuerpos de igual volumen. Un cubo de volumen ν tiene arista $\nu^{1/3}$ y una relación de superficie a volumen igual a $6(\nu^{1/3})^2/\nu$, es decir, $6/\nu^{1/3}$. Una esfera de este mismo volumen tiene diámetro igual a $(6\nu/\pi)^{1/3}$ y una relación de superficie a volumen de $\pi(6\nu/\pi)^{2/3}$, aproximadamente, $4{,}8/\nu^{1/3}$. El área de la superficie esférica, hace notar Jargocki, es aproximadamente un 20 por ciento inferior a la de un cubo de igual volumen. Un razonamiento semejante muestra que para un área superficial dada, el volumen de la esfera aventaja al del cubo en un 39 por ciento.

Tenemos una falacia muy semejante en el sorprendente teorema que sigue, enviado hace ya varios años por Thomas D. Waughn. Encerrando una esfera de radio *r* en un poliedro cuyas caras sean todas tangentes a la esfera, la relación de superficie a volumen es para el poliedro igual a $3/r$, independientemente de su forma, o del número de caras. Pero ¡también la relación de superficie a volumen de la esfera es $3/r$! Estas dos paradojas pueden servir para mostrar la necesidad de lo que en física se conoce por análisis dimensional, valiosa técnica para expresar constantes mediante números sin dimensiones, es decir, mediante números independientes de las unidades de medida que hayan podido adoptarse.

He comentado ya en varias ocasiones el problema de situar *n* superdamas, o amazonas (piezas que reúnen los movimientos de dama y caballo de ajedrez) en tableros de *n* por *n* sin que ningún par de ellas se amenace. El problema estaba sin resolver para todos los valores de *n,* pero ahora ha podido confirmarse la conjetura de que existe solución para todo valor de *n* mayor que 9. El año pasado Ashok K. Chandra, del Thomas J. Watson Research Center, de la International Business Machines Corporation, consiguió probarlo para todos los casos, excepto para $n = 20$, $n = 21$, $n = 32$, $n = 33$ y $n = 57$. Como ya dí a conocer el mes de julio, se han descubierto ahora soluciones para estos casos excepcionales, con lo que el problema está resuelto. A principios de año recibí otra solución general enviada por J. Reineke y P. Päppinghaus, de la Universidad de Hannover. Y en junio me llegó todavía otra solución más, de Charles Zimmermann, de Madison, Wisconsin. Tanto Zimmermann como Paul Steves, también de Madison, prepararon independientemente programas de ordenador capaces de determinar todas las soluciones básicas para tableros de orden bajo. Existe exactamente una solución para el caso $n = 10$, y hay seis en el caso $n = 11$, 22 para $n = 12$ (y no 23 como dije en alguna ocasión), 239 para $n = 13$ y 653 para $n = 14$.

Taller y laboratorio

Un láser casero de vapor de mercurio emite, a la vez, verde y anaranjado

Jearl Walker

Dean Morelli, de Rye, New York, cuyo espectrómetro de fabricación casera les describí en mi artículo de marzo pasado, ha ideado un procedimiento para que los aficionados puedan construirse un láser de vapor de mercurio ionizado. El láser de Morelli emite pulsos muy intensos de luz verde (de 567,7 nanómetros de longitud de onda) y otros pulsos más débiles de luz anaranjada (de 615 nanómetros). Este tipo de láser difiere de los otros en varios puntos. Se trata del primer láser de iones gaseosos y su ganancia y la anchura de su haz son desacostumbradamente grandes en un láser que emite luz visible. (Y, en realidad, fue con un láser de este tipo con el que se puso de manifiesto por primera vez la posibilidad de obtener ganancias y potencias elevadas dentro del espectro visible.) Sus emisiones próximas a los 615 nanómetros son casi monocromáticas puras, es decir, las frecuencias y longitudes de onda varían en dicha zona entre límites muy estrechos, por lo cual este aparato podría servir como patrón de frecuencias. En muchos aspectos, la construcción de un láser de Morelli es más fácil que la de un láser de helio-neón.

La ganancia de un láser constituye una medida de la amplificación de la intensidad luminosa por dicho haz (mediante la emisión estimulada) respecto a sus pérdidas (por dispersión, difracción y absorción). Para el verde, la ganancia del láser de Morelli está probablemente en torno al 50 por ciento por metro a lo largo del tubo del aparato. Entre los espejos de láser, la intensidad de la luz es muy elevada. Colocando el objeto a estudiar dentro de la cavidad del láser, donde está sometido a una luz muy intensa, podemos efectuar varios experimentos de espectroscopía. (La parte del haz que emerja del láser podrá incluso aprovecharse para otros experimentos.)

El tubo del láser de Morelli contiene vapor de mercurio. El láser se excita eléctricamente mediante una descarga de alta tensión a través del tubo. Cuando uno de los electrones de la descarga choca con un átomo de mercurio, éste pierde un electrón: se tiene entonces un átomo de mercurio ionizado en un nivel energético excitado. Los iones de mercurio pueden excitarse hasta distintos niveles energéticos, comenzando por los que caen espontáneamente a niveles más bajos, emitiendo luz cuando ello ocurre. Esta luz constituye la luminiscencia, pero no el haz, que desprende el tubo del láser. Parte de los iones estarán excitados en un nivel energético que es el significativo por lo que al funcionamiento del láser respecta (el nivel denominado $5f^2F^0{}_{7/2}$ en espectroscopía). Todo ion que se encuentre en dicho nivel puede emitir un fotón de una longitud de onda de 567,7 nanómetros si cae al nivel energético inferior $6d^2D_{5/2}$.

Algunos de los fotones emitidos de esta forma se mueven a lo largo del eje del tubo del láser. Cuando uno se cruza con otros iones de mercurio que estén en el nivel $5f^2F^0{}_{7/2}$, el campo eléctrico del primero puede estimular en los segundos la misma transición descendente. En cada transición se emite otro fotón, que es idéntico al anterior en fase, longitud de onda asociada y dirección de movimiento. Así, en el momento en que el primer fotón alcanza el extremo del tubo habrá estimulado la emisión de un gran número de fotones iguales a él.

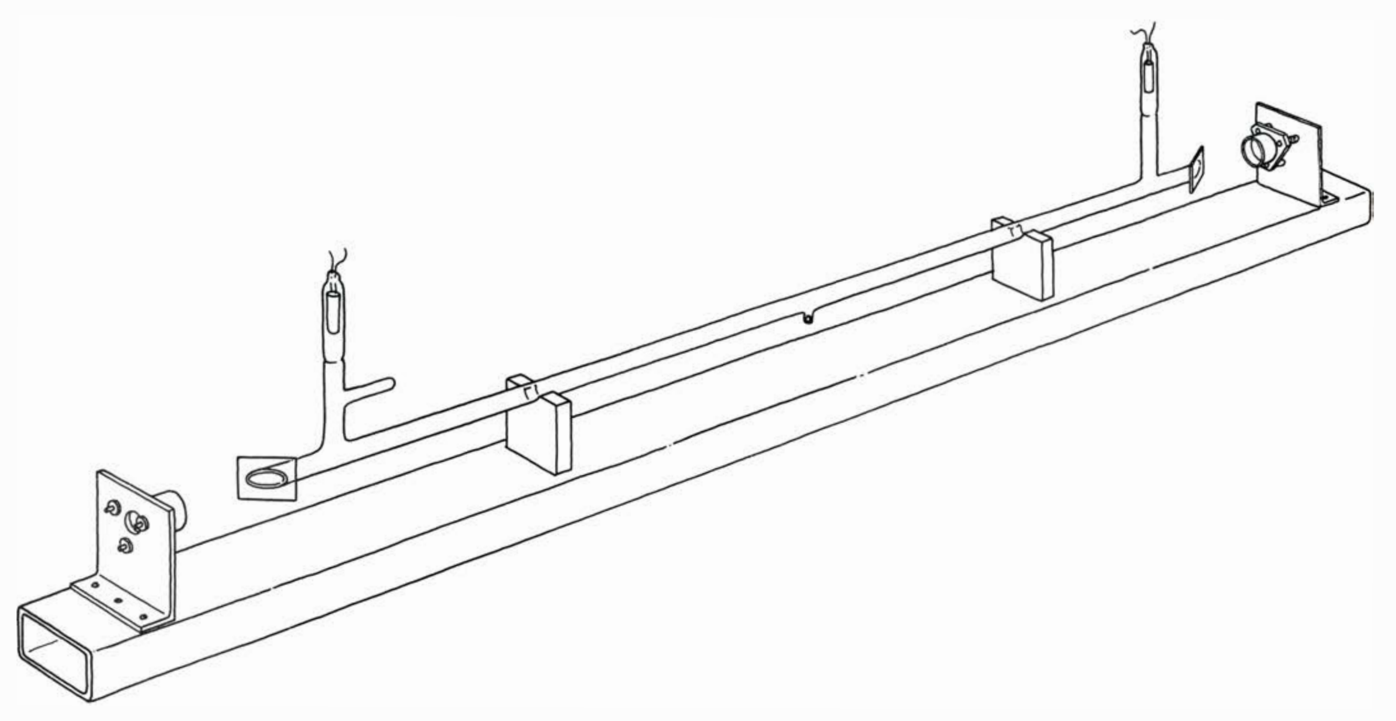
Llegados aquí, la descripción de la emisión estimulada suele hacerse más clara pasando de la representación de la luz mediante fotones a la descripción ondulatoria. Se dice que las ondas luminosas generadas en una emisión estimulada son coherentes porque se propagan en la misma dirección, tienen la misma longitud de onda y están en fase. Por tanto, se interfieren constructivamente, en virtud de lo cual se refuerzan entre sí y producen de esta suerte una onda de gran amplitud, lo que significa que la luz será muy intensa.

Las ondas luminosas se reflejan alternativamente, una y otra vez, a través del tubo de mercurio en los espejos colocados en sus extremos; ello estimula aún más las emisiones a la longitud de onda de 567,7 nanómetros. Parte de la luz atraviesa los espejos para formar el haz de láser exterior. ¿Qué ocurre con el resto de la luz emitida por los átomos de mercurio del tubo, tanto los no ionizados como los simplemente ionizados que no participan en la emisión estimulada? Cada átomo de éstos se halla en un nivel energético inadecuado o bien emite luz espontáneamente antes de que lo estimule un fotón que se cruce con él. La luz producida por estos átomos sale del tubo en todas direcciones y contribuye a su luminiscencia. Se trata de luz incoherente, lo que significa que las ondas emitidas no tienen todas la misma longitud de onda, ni se desplazan en la misma dirección ni están en fase.

La mayoría de los niveles energéticos alcanzados por los átomos de mercurio simplemente ionizados no participan en el efecto láser debido a que, por término medio, los iones permanecen en tales niveles un tiempo demasiado corto. Para que una determinada transición entre niveles energéticos contribuya al efecto láser, en el nivel superior debe haber más átomos que en el inferior. Esta situación se llama inversión de población.

Supongamos ausencia de inversión de población, de forma que, para un par de niveles, habrá más átomos en el inferior que en el superior. Aun así, la transición descendente fortuita de un átomo podría enviar un fotón a lo largo del tubo y este fotón podría producir una emisión estimulada al cruzarse con un ion de mercurio que estuviera en el nivel superior. Sin embargo, al no haber inversión, lo más probable es que los fotones estimulados encuentren átomos en el nivel inferior y no en el superior, y todo fotón que se cruce con un átomo en el nivel inferior será absorbido por éste. En ese caso, predomina la absorción y el resultado es la ausencia de efecto láser.

Para que sea posible la inversión de población ha de existir un nivel energético en el que los átomos puedan permanecer durante un tiempo largo en proporción a la duración de su vida en un nivel inferior hacia el que puedan caer. Los átomos del nivel inferior caen a niveles aún más bajos muy rápidamente, dejando vacío el nivel que ocuparon y relativamente lleno el más alto de todos; el resultado es la inversión de población.



Láser de vapor de mercurio ionizado construido por Dean Morelli

Electrodos de neon recubiertos; diámetro exterior, 15 mm.

15 cm. (aprox.)

Toma de vacío; long., 4 cm; diam. ext., 7 mm.

Ventana de Brewster de cuarzo soldada; dimensiones, 25 × 35 × 1 mm.

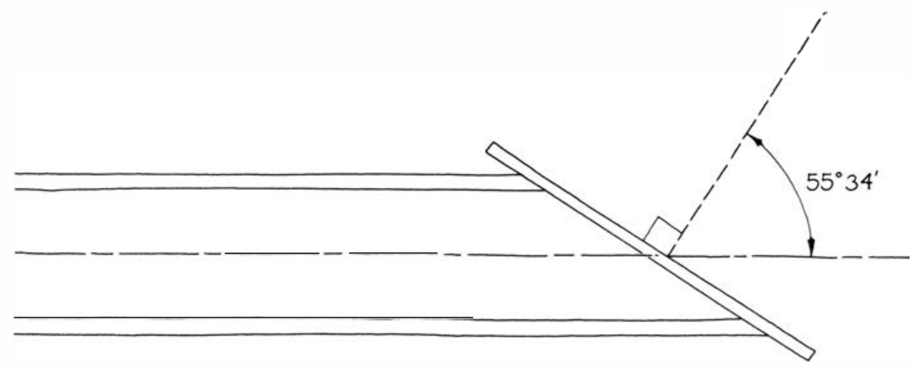
Tubo de 15 mm; espesor tipificado

Copa del mercurio; longitud, 2 ÷ 2,5 cm; diam. ext, 7 ÷ 8 mm.

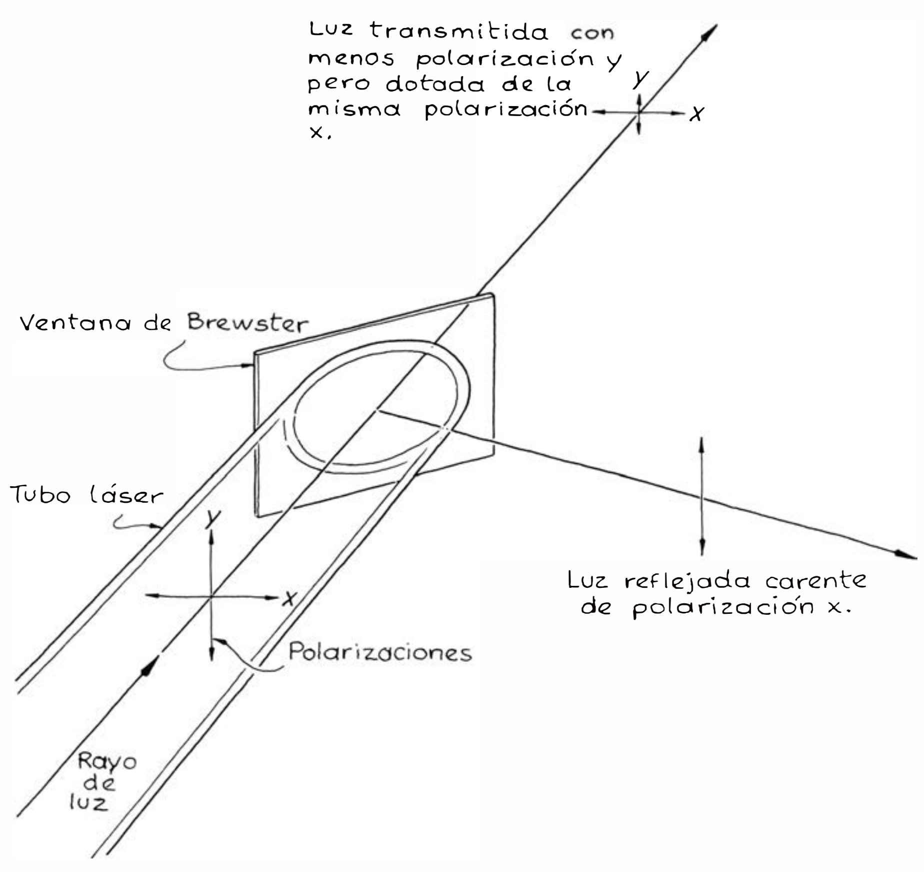
107 cm.

122 cm.

Características del tubo de descarga



Angulo de la ventana de Brewster



Efecto de la ventana de Brewster sobre la polarización

El nivel energético del ion de mercurio simplemente ionizado, capaz de originar luz láser de 615 nanómetros, es el denominado $7p^2P^0{}_{3/2}$ y la transición se hace al nivel $7s^2S_{1/2}$. Parte de la luz estimulada atraviesa un espejo para formar otro haz de láser que se encuentra en la región anaranjada del espectro visible.

En las condiciones correctas, el láser de Morelli emite ambas longitudes de onda (567,7 y 615 nanómetros) simultáneamente y los dos colores pueden separarse haciendo pasar el haz de salida del láser a través de un prisma. Sin embargo, la componente anaranjada es relativamente débil y por ello Morelli ha proyectado su láser pensando en la componente verde.

Morelli construyó el tubo de su láser a partir de una tubería de pirex de longitud tipificada (122 centímetros) y espesor tipificado, cuyo diámetro exterior es 15 milímetros. De otra tubería pirex cortó una longitud de 7,5 centímetros y la soldó a estanqueidad al tubo largo, a 7,5 centímetros de uno de los extremos de éste y formando ángulo recto. Para soldar empleó un soplete de oxígeno. Al extremo abierto del tubo corto fusionó un electrodo de lámpara de neón de 15 mm de diámetro. De la misma forma, montó otro electrodo en el extremo opuesto del tubo largo.

Seguidamente, a uno de los tubos de electrodo unió un trozo de tubería de 4 centímetros de longitud y 7 milímetros de diámetro exterior para que sirviese como toma de vacío. Otro trozo de este tubo de 7 milímetros lo fusionó al centro del tubo más largo por el lado opuesto a los electrodos. Este trozo, de 2,5 centímetros de longitud, constituiría la fuente de vapor de mercurio.

Si no se dispone de soplete de oxígeno, el montaje puede hacerse con tubería de vidrio sodocálcico, que puede fundirse con un soplete de propano como los que se venden en las ferreterías. En la bibliografía citaremos dos libros que resultan especialmente provechosos: *Creative Glass Blowing,* para los trabajos con vidrio, y *Light and Its Uses,* que ofrece indicaciones y consejos para la construcción de láseres.

Utilizando sendas ventanas de Brewster, construidas con vidrio plano de gran calidad óptica, Morelli cerró herméticamente los extremos del tubo largo. Estas dos piezas han de montarse sobre el tubo con el ángulo exacto para minimizar las pérdidas que sufre la luz al reflejarse en las ventanas. El ángulo es el formado por el eje del tubo y una recta perpendicular a la superficie de la ventana.

Este ángulo de Brewster, que así se llama, es igual al arco tangente del índice de refracción de la ventana y ejerce un influjo determinante sobre la polarización de la luz. Al comenzar el efecto láser, la luz incidente sobre la ventana no está polarizada; sin embargo, y merced al ángulo de Brewster, hay un sentido de polarización que resulta reforzado y que rápidamente llega a predominar en el proceso.

Supongamos un rayo de luz no polarizada que incida sobre una ventana de Brewster. Aquí el campo eléctrico luminoso oscilará a lo largo de todos los ejes posibles contenidos en un plano perpendicular al rayo. Resulta muy útil imaginar que las oscilaciones se efectúan a lo largo de dos ejes perpendiculares. Uno de ellos, el eje x, está en el plano definido por los rayos incidente y reflejado. El otro, eje y, es perpendicular al x. En razón de la geometría particular de la ventana de Brewster, la luz reflejada en su superficie está polarizada por completo en la dirección y. Esta luz sale del tubo al reflejarse y ya no contribuye al efecto láser. La luz polarizada según el eje x no se refleja en absoluto, por lo que atraviesa la ventana y es devuelta a través del tubo por uno de los espejos. De esta manera, la luz con polarización x llega a predominar en todo el efecto láser; la luz generada por emisión estimulada estará, pues, polarizada casi por completo, con lo que se evitará la reflexión en las ventanas y la luz que se pierda en éstas será muy poca.

Morelli empleó ventanas de cuarzo, cuyo índice de refracción para la longitud de onda de 567,7 nanómetros es 1,459. El arco tangente de 1,459 es 55 grados y 34 minutos y, por tanto, este fue el ángulo de montaje de las ventanas de Brewster. (La emisión de 615 nanómetros no llega a la ventana bajo el ángulo exactamente correcto, ya que la longitud de onda es más larga y le corresponde un índice de refracción ligeramente menor. No obstante, el ángulo de Brewster es casi igual para ambas longitudes de onda y, por ello, el láser puede trabajar a 615 nanómetros.)

Morelli preparó los extremos del tubo para recibir las ventanas aserrándolos a un ángulo complementario del de Brewster, o sea, 34 grados y 26 minutos. Este trabajo lo realizó con una sierra para metales dotada de una hoja de latón de 0,6 milímetros de grueso y manteniendo el tubo y un par de guías para la hoja de la sierra en una caja de ingletes. Como lubricante empleó abrasivo del número 120 mezclado con agua. Morelli recalca que la sierra debe moverse lenta y cuidadosamente para que no se astillen los bordes del vidrio, lo que originaría pérdidas de gas durante el funcionamiento del aparato.

Aserrado una mitad, se da la vuelta al tubo para terminar el resto del corte. Se pulen luego los bordes hasta dejarlos planos amolándolos contra un útil de vidriero alimentado sucesivamente con abrasivo cada vez más fino (el último y más fino de los cuales será del n.º 600). Los cortes no deben desviarse en más de medio grado del ángulo complementario del Brewster.

Finalizados el corte y el pulido, Morelli limpió el tubo y lo montó en su soporte. Con un cuentagotas introdujo entonces algunas gotas de mercurio en la copa, para lo cual inclinó el tubo ligeramente y con cuidado dejó que las gotas se deslizaran por el interior. Este mercurio ha de ser muy puro; él empleó mercurio tridestilado.

Una vez que el mercurio está en su sitio, se sumerge la copa en una cubeta llena de hielo seco y etanol hasta que el metal se congela. Después, la toma de vacío del tubo se conecta a una bomba de vacío. Con ésta en marcha, se sujetan las ventanas cuidadosamente contra los bordes del tubo. (Puede que aquí sea necesario un ayudante.) Si los extremos del tubo están suficientemente planos, la aspiración de la bomba sujetará firmemente las ventanas. A la menor fuga habrá que insistir en el pulido para que los bordes queden más planos. Si el contacto es hermético, las ventanas podrán encolarse en el tubo. Para ello, Morelli se sirvió de una jeringa hipodérmica desechable dotada de una aguja de calibre grande, con la que aplicó una capa de epoxia alrededor de las uniones de las ventanas con el tubo. La aspiración debe mantenerse hasta que la masilla se endurezca.

Como base para el tubo y los espejos empleó un trozo de tubería de aluminio rectangular de unos 183 centímetros de longitud, con una sección recta de 10 por 4,5 centímetros y 3 milímetros de espesor. Esta pieza es un trozo de tubería de las que se emplean en carpintería metálica. Como sostén del tubo utilizó dos soportes de madera cortados de una pieza de contraplacado de 2 centímetros de grueso y después aplicó epoxia entre los soportes y la base y entre los soportes y el tubo. (Para aplicar la epoxia al aluminio, Morelli preparó la superficie de éste restregándola con un cepillo de dientes empapado en un polvo limpiador abrasivo y prosiguió en este trabajo hasta que el agua depositada sobre el metal formaba una capa delgada sin tendencia alguna a globularse.)

Cada espejo se instaló en una montura ajustable mediante tres tuercas moleteadas. Para centrar los espejos con el tubo láser, Morelli siguió un procedimiento muy corriente en el que se emplean dos tarjetas de archivo, una pequeña lámpara portátil y un portaobjetos de microscopio que hace de separador de haz. En cada tarjeta practicó un agujero circular y fijó una a cada extremo del tubo, de suerte que los agujeros quedaran alineados con el tubo. El separador de haz lo colocó entre un espejo y el extremo del tubo correspondiente, formando un ángulo de 45 grados con el eje de éste. Además retiró la montura de este espejo con el fin de que pudiera observarse el espejo del extremo opuesto mirando a través del tubo.

Soporte de contraplacado

7.5 mm.

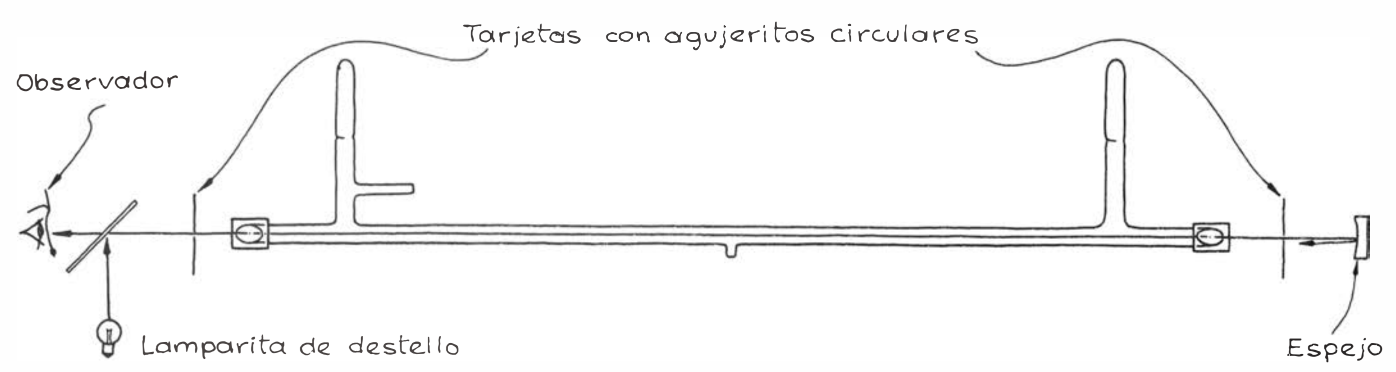
71 mm.

51 mm.

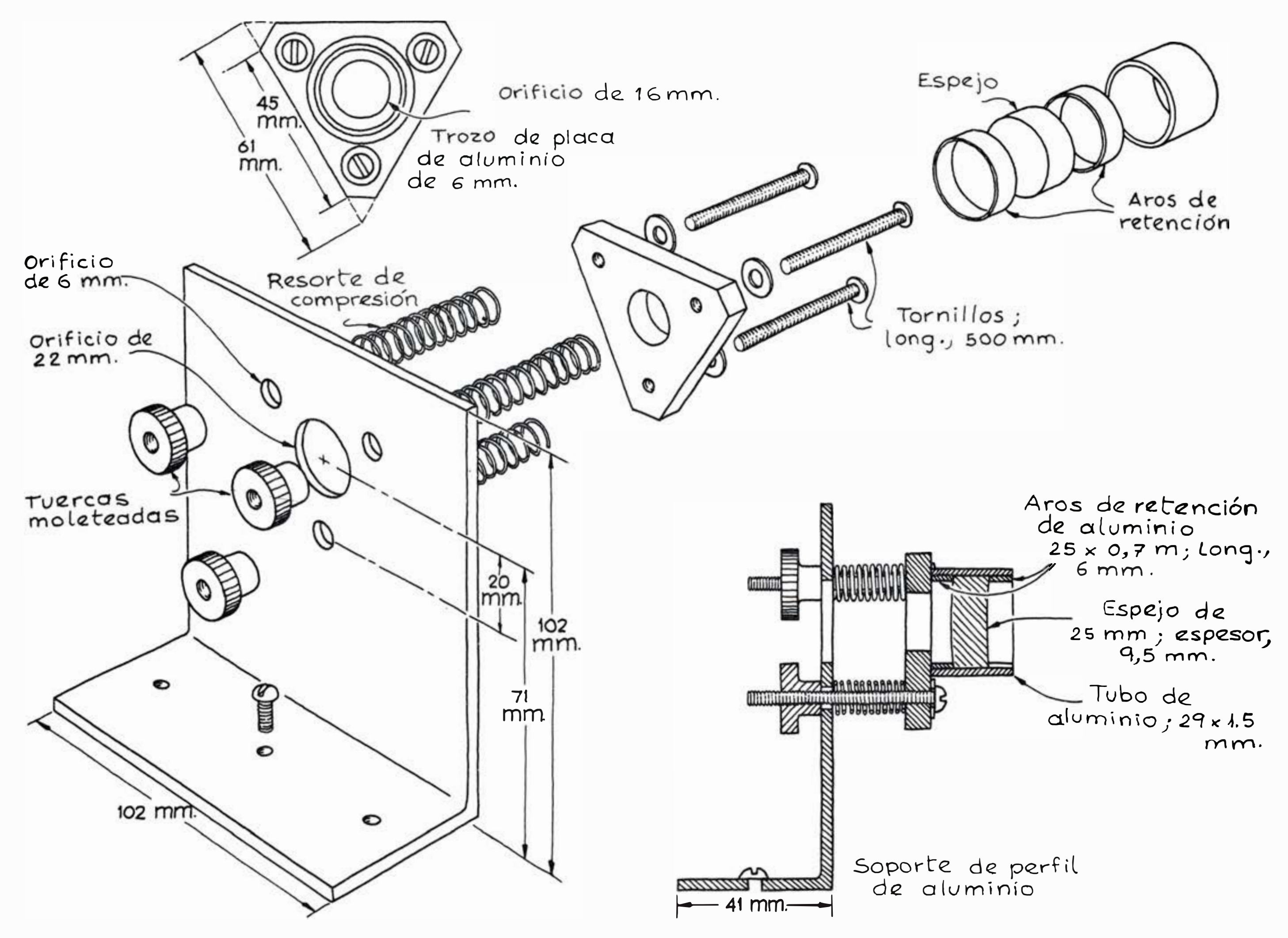
Soporte de madera para el tubo de descarga

Situó la lámpara, que debe tener un bulbo de filamento pequeño, junto al portaobjetos para que la luz se reflejase en éste y entrase en el tubo pasando por el agujero de la tarjeta más próxima. Ajustó, luego, los tornillos del espejo lejano (esto lo hará un ayudante, si se dispone de él) hasta que la luz se reflejó regresando al observador, cuya visual se dirigía a través del tubo. Obtuvo un centrado correcto cuando la intensidad de la luz procedente del espejo lejano era máxima.

Separó entonces esa montura de la base, sin alterar los tornillos de ajuste, de manera que al volver a instalarla siguiera estando ajustada. Morelli centró el otro espejo por el mismo procedimiento.



Procedimiento para centrar los espejos



Detalles de la montura de un espejo

Los espejos, esféricos, poseen el mismo radio; se colocan separados una distancia algo menor que un radio (disposición llamada cavidad confocal). Su radio de curvatura mide 147 centímetros, 25,4 el diámetro y 9,5 milímetros su espesor. El reverso de ambos está amolado en forma convexa. Su separación es del orden de 140 centímetros. (Las superficies exteriores podrían haber sido planas si se hubiera empleado una lente para colimar el haz emergente. Incluso hubieran podido ser planas las superficies interiores, pero los sistemas de espejos esféricos son más fáciles de centrar que los de espejos planos.)

Cada espejo se montó en un tubo de aluminio de 29 milímetros de longitud y 1,5 milímetros de espesor. El tubo se afirmó, con epoxia, a una placa triangular, y ésta se sujetó a una placa soporte con tres tornillos. En el interior del tubo se introdujo y se pegó en posición un aro de retención de aluminio de 25,4 milímetros de ancho por 6 milímetros de largo. Se introdujeron luego el espejo y otro aro. A fin de mantener en posición este aro, Morelli utilizó cinta transparente ya que, según hace notar, no conviene pegarlo, pues los espejos han de sacarse de vez en cuando para limpiarlos.

Los espejos que habitualmente se instalan en los láseres que emiten luz visible tienen varias películas dieléctricas delgadas pensadas para aumentar la reflectividad en una longitud de onda determinada. Los espejos empleados por Morelli estaban preparados para ofrecer la mejor reflexión a la longitud de onda de 567,7 nanómetros. Así, la película delgada contigua al vidrio tenía un índice de refracción superior al del vidrio y su anchura óptica era un cuarto de la longitud de onda designada.

Por encima de esta película hay otra capa de menor índice de refracción. Sobre éstas se añaden más capas, de índices de refracción alternativamente mayores y menores, con la exigencia de que la capa externa de cada pareja tenga el índice mayor. Cada una de estas capas refleja parte de la luz. Entonces, con este sistema, todos los rayos reflejados se interfieren constructivamente, proporcionando una reflexión intensa. Un espejo de éstos suele reflejar hasta el 99,8 por ciento de la luz que le llega, pero sin los recubrimientos la reflectividad sería considerablemente menor.

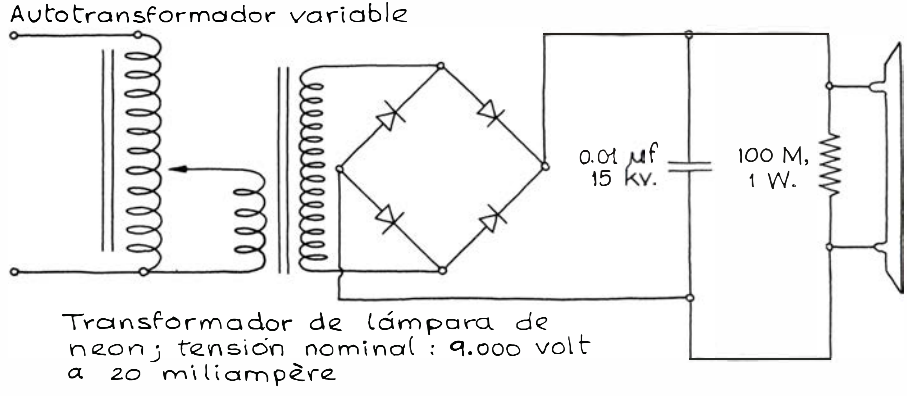
El espectro de longitudes de onda reflejado por un espejo depende del valor relativo de los índices de refracción de las películas de dieléctrico. Los espejos de Morelli reflejan aproximadamente el 99,5 por ciento de la luz para la longitud de onda de 567,7 nanómetros. El intervalo de reflectividad abarca aproximadamente unos 100 nanómetros de anchura espectral e incluye la emisión de 615 nanómetros que también puede observarse con este láser. No obstante, Morelli señala que, para esta longitud de onda, la reflexión es de bastante mala calidad, hasta el punto de que la emisión puede resultar muy débil o no aparecer.

El láser se activa mediante una descarga de alta tensión a través del tubo. Un autotransformador variable alimenta de corriente al arrollamiento

primario de un transformador de lámpara de neón de 9000 volt de tensión nominal (para 20 miliampère). La salida de este transformador se rectifica en un puente de cuatro diodos y luego se alimenta a un condensador en paralelo con el tubo. Cuando la diferencia de potencial entre los electrodos del tubo alcanza unos 5000 volt, el gas del tubo se ioniza; ello significa que puede conducir la electricidad, permitiendo que el condensador se descargue a través del tubo. Este arco de elevada intensidad dura alrededor de un microsegundo y, según sea el reglaje del autotransformador, la descarga puede repetirse con una frecuencia de hasta varios centenares de veces por segundo. Con una repetición tan rápida, la emisión de láser se diría que es un haz estacionario.

La capacidad del condensador debe ser de 0,01 microfarad aproximadamente y de 15 kilovolt su tensión de ruptura. Morelli lo construye a partir de un trozo de tarjeta de circuito impreso de dos caras. El área de la superficie de tarjeta necesaria (en metros cuadrados) está determinada por la fórmula $a = ct/d\epsilon_o$, donde ϵ_o es $8{,}85 \times 10^{-12}$ farad por metro, c representa la capacidad en farad, t simboliza el espesor del dieléctrico en metros y d es la constante dieléctrica del material aislante. La constante dieléctrica del material vítreo-epoxídico de la tarjeta de circuito impreso vale 4,8 y su tensión de ruptura es de 95 kilovolt por milímetro de espesor.

Por ataque químico, Morelli eliminó un centímetro del cobre existente a lo largo de ambas caras del perímetro de la tarjeta, con el fin de evitar la formación de arco en dicha zona. Luego tanteó experimentalmente para determinar cuál sería la mejor superficie de la tarjeta. Cuanto mayor es la superficie, mayor es la capacidad óptima. La capacidad es menor para la salida más intensa de 567,7 nanómetros que para el haz de 615 nanómetros.



Circuito de alimentación

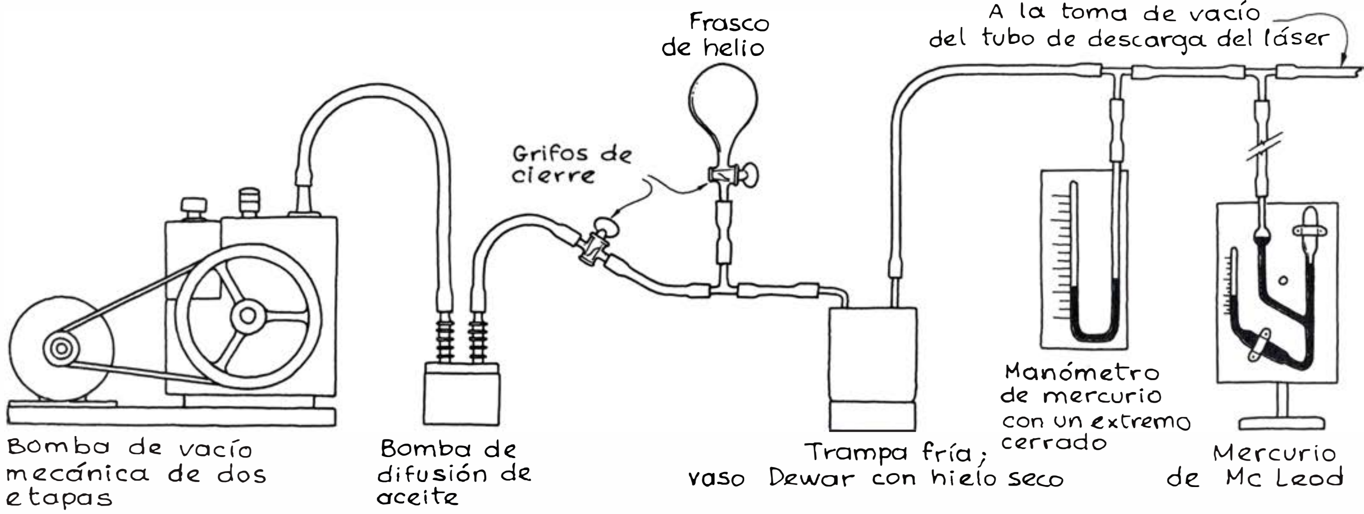
Los dos electrodos deben protegerse del bombardeo de electrones que reciben cuando el arco eléctrico discurre por el tubo. Morelli consiguió electrodos comerciales recubiertos de una mezcla de carbonato de estroncio y bario suspendida en un aglutinante de nitrocelulosa. Tras haber sellado el tubo del láser calentó los electrodos con el fin de convertir el recubrimiento en una capa protectora. Para el calentamiento se sirvió de una bobina introducida por el exterior del tubo del electrodo y excitada por un oscilador de alta frecuencia. La corriente oscilante de la bobina inducía una corriente oscilante en el electrodo. En paralelo con la bobina colocó un condensador variable para poner en resonancia el oscilador con el sistema condensador-bobina. La potencia nominal del oscilador debe ser de 50 watt como mínimo.

El calentamiento de los electrodos descompuso la nitrocelulosa, liberando hidrocarburos gaseosos, y los carbonatos se convirtieron en óxidos, liberando anhídrido carbónico. Los óxidos protegen los electrodos de la corriente de arco. Aquellas reacciones dieron además lugar a bario y estroncio metálicos que actúan como "degasificadores", eliminando del tubo gases activos e indeseables.

El circuito de vacío utilizado por Morelli consistía en una bomba de vacío mecánica de dos etapas, una bomba de difusión, un manómetro de mercurio de extremo cerrado, un manómetro de Mac Leod, una trampa fría, un frasco con helio y dos grifos de cierre. El sostiene que la bomba de difusión es útil, pero no imprescindible. Los planos de los manómetros se ofrecen en *Creative Glass Blowing*. Más instruciones acerca de éstos y de otros componentes del circuito de vacío aparecen en *Light and Its Uses* (léanse también los capítulos dedicados a los láseres de helio-neón y de argón).

Lo primero que hizo Morelli al poner en marcha el circuito de vacío fue congelar el mercurio contenido en la copa central. Arrancó después las bombas y comenzó a calentar el tubo del láser con un mechero Bunsen para eliminar las moléculas de gas adheridas a sus paredes, poniendo buen cuidado en no calentar las ventanas de Brewster ni el mercurio contenido en la copa.

Una vez preparados los electrodos



Circuito de vacío

con su recubrimiento de óxido, Morelli los conectó a la salida de corriente alterna del transformador de lámpara de neón. Después introdujo helio en el tubo a la presión de algunos torr, con lo que se formó una descarga luminiscente que limpió el interior del tubo. Este helio se bombeó al exterior y se introdujo helio nuevo, repitiéndose el proceso hasta que la descarga generó una luminiscencia rosa. Entonces se ajustó la presión del helio entre 0,5 y un torr, se introdujo la copa de mercurio en una cubeta con agua a 40 grados centígrados y se conectaron los electrodos a la fuente de alimentación de pulsos.

Una vez que el mercurio se haya difundido por el interior del tubo, se ajustan los espejos hasta que el láser comience a funcionar, al menos en la región del verde. Para que la salida sea máxima deben hacerse varios ajustes. Las monturas de los espejos pueden necesitar otros ajustes más. Por su parte, el operador puede hacer variar la presión del gas, la temperatura de la copa del mercurio y la capacidad del condensador de la fuente de alimentación.

Morelli sugiere que periódicamente se inviertan los cables de los electrodos, de manera que éstos intercambien sus papeles como ánodo y cátodo. Con esto se evita que sea uno de los electrodos el que reciba siempre el bombardeo de electrones y, además, que el helio se impulse únicamente hacia uno de los extremos del tubo. La presión del gas disminuirá gradualmente a medida que parte del helio quede sepultado por material de los electrodos pulverizado. Tras varias horas de funcionamiento hay que reponer el helio.

No voy a extenderme sobre la peligrosidad potencial de los láseres. Casi todo el mundo sabe que, si un haz de láser penetra en el ojo, puede causar una lesión permanente en la retina; pero el peligro se encierra también en las reflexiones esporádicas del haz sobre objetos de la habitación. Hay que andar con cuidado con las corrientes de elevada intensidad que se desarrollan en la fuente de alimentación, peligro que se agudiza especialmente si se trabaja con el láser en una oscuridad parcial que impida la visión completa del circuito. Hay que aislar todos los cables exteriores. Préstese atención también al manejo del mercurio, pues éste no sólo se evapora a altas temperaturas, sino también a la del recinto, y su vapor es venenoso. Por ello, al cargar el mercurio en la copa se comprobará previamente que la habitación esté bien ventilada.

Libros

Dos libros clásicos de suerte parecida

Ramón Margalef y Silverio Palafox

AMBIENTE, ENERGÍA Y SOCIEDAD, por Howard T. Odum. Ed. Blume, Barcelona, 1980; 409 págs. Los hermanos Eugene P. y Howard T. Odum son muy conocidos por haber contribuido de manera distinguida a la configuración que ha ido tomando la ecología moderna; el primero es autor de un manual de ecología, muy difundido también en su versión latinoamericana. Howard (Tom) colaboró en el libro mencionado; pero su aportación a la ciencia ecológica ha sido menos convencional, consistente más bien en la continua exploración de nuevas vías de acceso a la comprensión de la naturaleza, con imaginación desbordante, siempre estimulante y casi siempre creadora. Recomiendo encarecidamente el libro cuya traducción nos ofrece ahora Blume, dentro de su serie ecológica. Publicado por primera vez en 1971, en Estados Unidos, con el título "Environment, power and society", podría parecer algo añejo; pero, a diferencia de tantos libros más o menos ecológicos, que pasan sin pena ni gloria, éste es de valor perdurable y tuvo y tiene el mérito de anticipar y presentar, en un contexto racional, algunas de las consecuencias previsibles cuando se limita el acceso de nuestra sociedad a la energía fósil. La energía es el hilo conductor que unifica el contenido del libro, que explora la posibilidad de usarla como patrón común en el estudio comparado, no solo de sistemas propiamente ecológicos, sino también de los económicos y sociales. Por supuesto, la energía informa una red activa de conexiones que vendría a ser la esencia de todos estos sistemas, cuyo estudio queda así dentro de una teoría general de sistemas. En este libro, H. T. Odum usa ampliamente el simbolismo que ha inventado para caracterizar los diferentes elementos de un sistema, según su función de almacenamiento, conmutación, refuerzo, etc. Los diagramas resultantes, una vez entendidos, clarifican considerablemente la discusión de los sistemas más complicados. Describimos brevemente a continuación el contenido del libro, con indicación de algunos de los temas que personalmente creo de mayor interés.

El capítulo 1 (El sistema de este mundo, 29 págs.) destaca la importancia de la energía, también de la externa o exosomática, tradicionalmente tan descuidada por el naturalista, y sugiere como puede ser más importante que la virtud o la ideología en configurar la sociedad humana y determinar sus cambios; la teoría de sistemas se aplica a la biosfera. El capítulo 2 (Qué es la potencia, 38 págs.) nos encara con el problema secundario de traducir "power". El traductor prefiere potencia, o capacidad de realizar trabajo; en la forma en que generalmente se usa energía, casi sería preferible adoptar esta palabra. El flujo de energía en sistemas con los que estamos familiarizados sirve de introducción. Se exponen los rudimentos de la termodinámica, se trata de la reversibilidad, optimación, relaciones entre flujos y fuerzas y se introducen los símbolos que el autor utiliza en sus esquemas. (Hay un apéndice al final del libro que compendia las fórmulas y los símbolos de los elementos componentes.)

El tercer capítulo (La potencia en los sistemas ecológicos; 54 págs.) constituye un excelente resumen de ecología, basado en la descripción de la organización de las redes tróficas, sus flujos y eficiencia. El análisis de la organización de los ecosistemas terrestres y acuáticos a lo largo del eje vertical es un prodigio de síntesis. Diferencias horizontales, las relaciones entre producción y biomasa, fluctuaciones más o menos periódicas y la diversidad, se tratan a continuación, terminando con la consideración de los ecosistemas mundiales. En el capítulo 4 (La base de la potencia para el hombre, 45 págs.) se empieza por considerar el mantenimiento de poblaciones humanas primitivas en ecosistemas "naturales", pasando a analizar los circuitos de relaciones, en términos de energía, para las distintas civilizaciones, las consecuencias de los movimientos individuales sobre el plano horizontal y el subsidio de energía externa. En efecto, el contenido energético de nuestros alimentos procede ahora en parte del petróleo. El capítulo 5 (Potencia para el orden y la evolución; 42 págs.) trata de la adquisición y mantenimiento del orden en la vida, como contrapartida del aumento de entropía en un sistema más extenso. Concilia a Schrödinger y a Darwin. La selección natural conduce al desarrollo de bucles de refuerzo o de recompensa en extensas áreas de los sistemas naturales. Es interesante la discusión del origen de la vida y el papel que el autor asigna a los circuitos energéticos, preeminentes sobre los puntos de vista más corrientes que siempre piensan en piezas prefabricadas. En el capítulo 6 (Potencia y economía; 37 páginas) se exploran las relaciones entre cualquier representación de los sistemas ecológicos y las matrices de entradas y salidas que usan los economistas. Numerosos conceptos y problemas económicos se sitúan –a veces de modo algo forzado– dentro del mismo esquema general. Considero muy interesante su forma de tratar el dinero, como un componente producido por un bucle de retroacción que se inicia en otra parte de la sociedad y fluye en sentido opuesto a la energía, pero que no es una medida exacta de la energía. En esto estoy de acuerdo, pero me hubiera gustado hallar una exploración más profunda del significado del dinero como capacidad de anticipación, concepto que no sé si es original, pero facilita una interpretación ecológica y, de pasada, explica la inflación. Los avances de la medicina eliminan circuitos de retroacción biológicos. La ley de los rendimentos decrecientes se relaciona con las expresiones del tipo de Michaelis-Menten. Se consideran las relaciones entre países y grupos de países desarrollados y en desarrollo; pero no acepta explícitamente la fatalidad del principio de San Mateo.

El capítulo 7 (Potencia y política; 37 págs.) considera que la política consiste en un control superpuesto a los flujos de energía y traza los circuitos correspondientes a la dictadura y a la democracia. Al aumentar el número de unidades, disminuye la eficacia, según la conocida ley de Parkinson. La libertad individual siempre sale mermada cuando el conjunto funciona como un sistema. Se pregunta qué es más eficiente, estados pequeños que han alcanzado fronteras más o menos estables a través de pequeñas guerras periódicas, o una enorme organización

mundial, y cuál es la mejor forma de desarrollo, si por organización a partir de una situación anárquica inicial o si es posible trasplantar una organización desde fuera (el referente se siente tentado a terciar también en esta discusión, diciendo que apenas cree en la eficacia de los trasplantes de organización, lo que no desea sea interpretado como una apología del anarquismo). Después de considerar problemas de derecho, de la distribución de la tierra y muchas otras cuestiones interesantes, se pregunta, tras ilustres antecesores, si es posible una teoría de la historia, si las civilizaciones reproducen algunas características de la sucesión ecológica e inquiere audazmente si no sería hora de que el historiador utilizara el ordenador para simular los futuros. Algunas de estas ideas se comentan con referencia especial al Imperio Romano, lo cual desvía un poco la urgencia de los problemas que tenemos ahora y aquí con nosotros. El capítulo 8 (Base energética de la religión; 23 páginas) completa este itinerario, tratando de ver si es posible considerar a la moralidad como un conjunto de ciertas normas de programación de circuitos de energía, y preguntándose si su valor de supervivencia explicaría la adecuación de ciertas formas de religión para cada momento. El dios del pasado se reconoce en la supervivencia, y continuamente aparecen falsos ídolos, entre los que específicamente nombra al automóvil, como perturbadores de una red armónica de relaciones que está fundamentada en la energía.

En el capítulo 9 (Simulación eléctrica de redes energéticas; 23 págs.) se trata de hacer modelos de ecosistemas, principalmente modelos análogos, que pueden consistir en circuitos eléctricos. Puede ser muy educativo para los que se han orientado exclusivamente hacia modelos basados en cálculo digital. Se consideran en aquella versión los modelos de Volterra y Lotka. Es interesante el estudio de cómo se propagan y alteran las pulsaciones y la consideración del ecosistema como si fuera su propio ordenador. El capítulo 10 trata, en 37 páginas, de la asociación ("Partnership" en el original, que no es exactamente lo mismo) del hombre con la naturaleza, y contrasta la situación en la antigüedad, cuando el hombre era un pequeño bucle en el ecosistema, con la gran capacidad de control que actualmente posee. Todos los problemas de contaminación, perturbaciones diversas y explotación son más difíciles de abordar de lo que parece. Comenta la pesadilla que sería para un técnico electrónico asistir a la constante multiplicación y evolución de los elementos de sus circuitos, que es lo que ocurre con los sistemas de la biosfera. (La conclusión es que las pretensiones de los ecólogos modeladores son excesivas.) Los ecosistemas cerrados para mantenimiento de la vida, y para el reciclado, tal como se piensa (o se pensaba en 1971) que podrían funcionar en estaciones del espacio, constituyen un buen tema de discusión. Muchas estructuras son valiosas por constituir bucles de refuerzo o de estímulo para determinadas vías de reciclado. El capítulo final (11) se pregunta ¿Qué sistemas vendrán después? (8 págs.) y examina las distintas perspectivas, según que las disponibilidades de energía aumenten, permanezcan constantes o disminuyan. Ve que la competencia entre grupos humanos no depende de la población, sino de la energía que manejan (uno puede preguntarse hasta qué punto y en qué medida esto depende de la estabilidad social), y que el control en el punto donde se obtiene la energía puede ser decisivo. Como es general entre naturalistas no ecologistas, prevé una reducción en el ritmo de expansión, o incluso un decrecimiento, pero sin un colapso brusco o catastrófico. Es optimista, porque opina que probablemente saldremos mejor librados que en la transición del Imperio Romano a la Edad Media. Hace falta una nueva ingeniería, pero no basta la técnica, sino que es más importante la motivación y quizá deberemos resucitar modelos antiguos de sociedad.

Creo que este libro es de un nivel intelectual superior al que nos tiene acostumbrados la literatura ecológica, y puede educar a un público bastante amplio. Probablemente será poco leído, en parte porque las dificultades de la obra no han sido precisamente allanadas por la traducción; muchos párrafos con notable inspiración y hasta lirismo en el original quedan más bien chafados en castellano; los epígrafes de las tablas a veces son poco claros, y hubiera sido aconsejable utilizar unidades métricas en todo el libro. Una de las pocas veces que se dan equivalencias, por errata de imprenta se pone (p. 66) 1 galón = 4546 litros. Caloría gramo sería más apropiado que gramos caloría. Un programa (pág. 333) podría expresarse en términos más de acuerdo con el uso internacional, que no requiere una traducción tan literal. Aunque este libro me gusta mucho, me pregunto por qué se ha traducido en vez de "Energy basis for man and nature" del propio Howard T. Odum, en colaboración con su esposa E. C. Odum, aparecido en 1976 (McGraw-Hill), que contiene sustancialmente las mismas ideas, pero presentadas de forma mucho más sencilla y asequible, está dirigido a un público más amplio, es muy apropiado para las escuelas y quizá hubiera resultado de mayor actualidad al acusar el impacto directo de los primeros pellizcos de la escasez, más o menos artificiosa, pero indudablemente oportuna, del petróleo.

Sin embargo, el libro de 1971, más extenso, puede ser de más provecho para el lector realmente preocupado. Repito que el tema básico es que la energía es un concepto unificador en todo el estudio de los sistemas naturales y humanizados. La importancia de la energía externa se ilustra excelentemente en la falacia de considerar que los cultivos de algas unicelulares puedan ser rentables; el autor tiene sus dudas sobre el futuro de la energía nuclear, que por ahora requiere un enorme subsidio de energía externa de otro origen. Las características de nuestra agricultura, de nuestra ganadería, dependen del subsidio de energía externa, lo que no siempre se tiene en cuenta en pretendidas transferencias de tecnología. Los consejos que los países desarrollados dan a los poco desarrollados toman, en consecuencia, un cariz tragicómico. Aunque la biomasa y la información quedan un poco relegadas a segundo término, su importancia se destaca donde conviene: las formas muy complicadas del comportamiento animal actúan como programas de control de los sistemas. Señales de colores, que establecen relaciones significativas, por ejemplo, entre insectos y aves, son casi comparables a los hilos de un circuito codificados con distintos colores. La ciencia entra dentro del mismo esquema, sólo se conservan aquellas ideas que se refuerzan por un bucle o circuito de validación empírica. Muchas de las críticas y comentarios sobre el orden actual que aparecen aquí y allá en el texto son muy pertinentes. Se comentan las respectivas ventajas e inconvenientes de los sistemas o países grandes o pequeños, también desde un punto de vista que muchos autores, y entre ellos los más entusiastas, descuidan; a saber, que no es posible desconectar de la energía externa disponible el tamaño del "grano" o "celdilla" unidad de organización. ¿Vale la pena seguir dedicándonos al cultivo de viejos tejidos humanos? ¿Qué energía se invierte en el control de la natalidad? Pero el autor no explicita que éste sea un flagrante ejemplo de un circuito retroactivo positivo de mantenimiento de información. El viejo, más por viejo que por otra cosa, llega a dominar las

artimañas para seguir viviendo y la gerontocracia es esperable. El autor no es desarrollista. Se pregunta ¿qué derecho tienen las compañías petroleras a explotar el crudo de Alaska? Pero no se hace muchas ilusiones acerca de la posibilidad de alcanzar cierto consenso por vía racional. Su capítulo sobre la religión ha de interpretarse como expresión de su convencimiento de que hace falta un refuerzo de algún otro tipo para armonizar o hacer menos conflictivo el comportamiento dentro de los grupos humanos, y el libro acaba preguntando enfáticamente: Profeta, ¿dónde estás? Una actitud humana y afirmativa que no está necesariamente reñida con la ciencia tradicional. (R. M.)

ANATOMÍA DEL SISTEMA NERVIOSO CENTRAL: TEXTO PROGRAMADO, por Richard L. Sidman y Murray Sidman. Ed. Fontanella; Barcelona, 1978. Libro de curiosa peripecia, esta "Neuroanatomía" programada de los Sidman (cuyo título comienza por cambiarse sin ventaja alguna al aparecer traducido al español) se vertió muy pronto al alemán, y se lo disputaron sin tardar editores en otras lenguas por la triple razón de que nada hay semejante en su nivel y de que está hecho por un lado con una sencillez franciscana y por otro con una calidad didáctica digna de toda alabanza.

Me precipito adrede en aclarar que el panegírico precedente vale y se queda corto para la edición original inglesa –de hace ya quince años– y para la inmediata traducción alemana, que conozco. Porque de la malhadada versión castellana no puede, infortunadamente, decirse nada similar, como podrá leerse enseguida en esta breve reseña.

Antes de entrar en ella, conviene aclarar algo más –que en la traducción ni se alude– respecto al plan de los autores, que, lamentablemente no han cumplido. En efecto, lo traducido y único publicado no es sino el primer tomo de una obra que para ser completa debería tener un segundo volumen. El primero comprende una inmejorable introducción –paso a paso, por items perfectamente diseñados y contrastados según la mejor técnica del aprendizaje programado– al estudio estructofuncional de la superficie cerebral, del sistema córticoespinal, tanto en sus relaciones con los núcleos basales como en su recorrido hasta la médula, así como la sistematización de ésta y del tronco cerebral, en sus estructuras propias y en cuanto se refiere al paso por ellas de vías somatoestésicas y vegetativas. El volumen segundo y último se anunciaba como una programación similar relativa al cerebelo y a las vías extrapiramidales, a las de la sensorialidad especial, y a la sistematización del diencéfalo y el córtex, al complejo ventricular, la anatomía vascular y otros complementos. No debo omitir que en ambos volúmenes se atendería –y en el primero así se hace– a los más esenciales aspectos fisiológicos y fisiopatológicos, al hilo de la exposición.

Como es patente, si interesante resultaba el primer tomo, más lo parecía aún el segundo, dado el total acierto del anterior. Cabe afirmar que, con

ambos, la obra completa de "Sidman and Sidman" sería perfecta, una verdadera joya de la bibliografía programada, y digna de una difusión sin límites en cualquier idioma. Con toda urgencia e interés se pidió por ello el segundo volumen apenas conocido el primero. Los telex de respuesta al importador fueron retrasando sucesivamente la probable fecha de su aparición, hasta que las respuestas se limitaron al consabido "no publicado", y en el actual catálogo de la editorial *Little, Brown and Company* ni se menciona ya el proyecto de tal segundo tomo, dejación más que lamentable, dado que su retardo en aparecer nunca le restaría un ápice de interés.

La versión española se veía por tanto, ya de entrada, limitada por no sabemos qué nefasto maleficio a la excelente pero incompleta primera parte de una obra inacabada. Pero añadiendo por su cuenta defectos en todo evitables, desde el título hasta el precio, se ha reducido en España el adecuadísimo tamaño del original, con lo que se emborronan no pocos de los clarísimos dibujos, inutilizados así para su didáctica finalidad, y se acompañan de un texto tan mal vertido, asintáctico, desorientador y hasta contradictorio con las más elementales nociones de la neuroanatomía funcional, que hace de su publicación, en vez de un celebrable evento editorial, un bochorno bibliográfico lamentable; lamentabilísimo por doquiera que se mire. Y dada la categoría original del libro, no hay más remedio –claro está– que considerarlo.

Como a nada llevaría entretenerse en detalles, vayan tan sólo unas pruebas de las gravísimas acusaciones precedentes, por lo demás al alcance de comprobar por cualquiera, pues abundan tanto los errores de todo tipo y calibre, que vale del todo el tópico: nada más fácil que multiplicar los ejemplos aducidos. Desde la gramática más elemental ("en el nivel *más superior*" = C459) hasta las faltas de concordancia sintáctica más increíbles ("y se integran *al* del lado"... "*izquierda*" = C460 y C460A); desde las imprecisiones más negligentes ("señale" por circunde; "del" por procedentes del = C462, *passim*) a las más desorientadoras ("surgen" en vez de proceden = C461, entre muchísimas equivalentes), llegando –a través de un auténtico galimatías sobre si ciertas fibras vienen de, van a, pasan por, etc., etc.– a reiteradísimos disparates y contradicciones (valga el de poner "suben" en vez de bajan tras haber escrito que son fibras descendentes... = C465). Todo hace ver que el traductor no tiene ni remota idea de por dónde se anda. Y la gravedad es máxima, porque la elementabilidad y la claridad del original también lo son; no cabe, pues, aducir atenuante alguna. Si he tomado números sucesivos es para que se advierta sin necesidad de extenderme que no se trata de erratas o inadvertencias sueltas. No hay página que se salve, cuando no por fas, es por nefas.

Así (otro grupito cercano valdrá por cientos), unas veces se traduce literal donde tal vez fuera mejor interpretar el sentido ("pobremente" por poco o escasamente = B242) y otras se cambia lo incambiable ("*el* ventrículo lateral", por un ventrículo lateral = B255). O bien –y es igualmente un sólo ejemplo– ¿por qué se cambia la redacción del B257, si no hace falta y es mejor la original, y sin embargo en B258, donde un pequeño cambio de perífrasis y preposiciones lo mejoraría mucho, no se hace? ¿Por qué en este mismo item no se pone el artículo (el, la), que es orientadoramente necesario en castellano, aunque no lo sea, por invariable, en inglés?

Modelo de incongruencias no menores es nombrar cada vez de una manera diversa las distintas regiones de la cápsula interna, con denominaciones frecuentemente inadecuadísimas, pues no es, en efecto, lo mismo miembro que extremidad, ni extremidad que extremo, a más de que en el original siempre emplea "limb" (B233A, por ejemplo). "*Et sic de coeteris*"... hasta la desesperación del lector más paciente y tolerante.

Naturalmente, ante tamaños disparatones nadie podía esperar, y por supuesto no se hace, ninguna mejora de los levísimos defectillos que, una vez impreso, cualquier estudioso del original podría encontrar en él. Uno evidente es, por único ejemplo, que el ítem 462 fuese el 461 y éste aquél. Pero finuras de este tipo que cabrían es claro que no se pueden pedir a quien se limita a fingir una seudotraducción.

No hay derecho; literalmente, no hay derecho a semejantes atropellos. Quizá podríamos haberlos temido, dado que la misma editorial publicó otro crimen de lesa traición seudotraductora hace ya años con la "Psicofisiología" de D. P. Kimble, francamente buena como iniciación elemental en su original "Physiological Psichology", aparecida casi diez años antes, por lo que no sólo se dio un texto anticuado sin la fácil actualización conveniente, sino que se imprimió un indecoroso galimatías seudocastellano, tan distorsionante e inservible para el neófito como innecesario para el ya iniciado. El problema que plantean estas versiones fictivas de buenas obras didácticas es realmente gravísimo, y deberían planteárselo los propios editores, pues aparte de dar la falsa impresión de que aquí no se sabe traducir (cosa que, para mayor INRI analfabético en textos programados como los que se comentan, es casi imposible por su intrínseca facilidad léxica), no sólo no sirven para iniciar sino que hacen creer el tema más difícil todavía de lo que en realidad pueda ser al lector que, enfrentado con un texto supuestamente excepcional para facilitar la iniciación, no logra introducirse en ella con cierta seguridad, y, por si todo ello fuese poco, privan nada menos que a todo el sector de los que leen en nuestra lengua del recurso de tal obra, puesto que con los derechos ya concedidos ningún otro puede enderezar el entuerto publicando una verdadera versión, por culpa de una tal falsedad, que para postre difamará en nuestro ámbito idiomático a sus autores y editores originarios, cargándoles un sambenito calumnioso, del que lo más probable es que ni siquiera lleguen a enterarse.

En suma, una posible gran obra, doblemente frustrada: *i*) ya en la edición original, porque le era imprescindible su segundo tomo, al que parecen haber renunciado; *ii*) en la versión española, porque ese primer tomo, aunque incompleto, era excelente, y entre traductor y editores lo han triturado literal, material y hasta económicamente (aunque esto sería lo de menos, si fuese lo único), puesto que se trataba de una obra intencionadamente sencilla para que no saliera más cara de lo necesario, a fin de que cada cual se la "machacara" personalmente hasta dominarla, casi como un cuaderno de prácticas. De recomendarla con entusiasmo al tener noticia de su próxima edición española, a tener que desaconsejarla con mayor y muy entristecido énfasis al contemplar lo que han hecho con ella, todo ha sido una.

Sólo me resta el consuelo pírrico de que no sea esta la primera recensión que publico en INVESTIGACIÓN Y CIENCIA, y que la anterior fuese tan loatoria como aquel libro merecía, para que el lector no me juzgue como un censor negativista. Y si bien se mira, una vez hecho el tan desagradable como obligado trabajo de redactar ésta, parece que podría volverse quizá tan eficaz como las más positivas, caso de valer como advertencia. Lo dudo, empero, tanto como lo deseo. (S. P.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

FISICA Y TECNOLOGIA DEL REACTOR DE FUSION

ASPEKTE DER ENERGIEVERSORGUNG MIT FUSIONSREAKTOREN. Bünde et al. en *Brennstoff-Wärme-Kraft,* vol. 26, n.° 11, 1974.

A FUSION POWER PLANT. Dirigido por R. G. Mills. Princeton Plasma Physics Laboratory, MATT-1050, 1974.

THE TECHNOLOGICAL REQUIREMENTS FOR POWER BY FUSION. D. Steiner en Proceedings of the IEEE, vol. 63, n.° 11; noviembre, 1975.

CONCEPTUAL DESIGN OF A TOKAMAK REACTOR. KFA Jülich. Technical Commitee Meeting and Workshop on Fusion Reactor Design, 10-21 de octubre de 1977. Universidad de Wisconsin, Madison, Estados Unidos.

EL CURIOSO ESPECTRO DE SS 433

A KINEMATIC MODEL FOR SS 433. George O. Abell y Bruce Margon en *Nature,* vol. 279, n.° 5715, págs. 701-703; 21 de junio de 1979.

ENORMOUS PERIODIC DOPPLER SHIFTS IN SS 433. Bruce Margon, Holland C. Ford, Steven A. Grandi y Remington P. S. Stone en *Astrophysical Journal Letters,* vol. 233, n.° 2, 2.ª parte, págs. L63-L68; 15 de octubre de 1979.

DOES ANYONE UNDERSTAND SS 433? Dennis Overbye en *Sky and Telescope,* vol. 58, n.° 6, págs. 510-516; diciembre, 1979.

ANTICUERPOS MONOCLONALES

LYMPHOCYTE HYBRIDOMAS. Dirigido por F. Melchers, Basel M. Potter y N. Warner. *Current Topics in Microbiology and Immunology,* vol. 81, 1978.

HYBRID MYELOMA MONOCLONAL ANTIBODIES AGAINST MHC PRODUCTS. Dirigido por Göran Möller. *Immunological Reviews,* vol. 47, 1979.

MONOCLONAL ANTIBODIES AND CELL SURFACE ANTIGENS. C. Milstein, G. Galfré, D. S. Secher y T. Springer en *Ciba Foundation Symposia,* n.° 66, págs. 251-276; 1979.

LAS CAUSAS DEL COLOR

THE ORIGINS OF COLOR IN MINERALS AND GEMS. K. Nassau en *Lapidary Journal,* vol. 29, págs. 920-928, agosto; págs. 1060-1070, septiembre; págs. 1250-1258, octubre y página 1521, noviembre, 1975.

THE ORIGINS OF COLOR IN MINERALS. Kurt Nassau en *The American Mineralogist,* vol. 63, n.° 3/4, págs. 219-229; marzo/abril, 1978.

LOS APALACHES MERIDIONALES Y EL CRECIMIENTO DE LOS CONTINENTES

LITHOSPHERE PLATE: CONTINENTAL MARGIN TECTONICS AND THE EVOLUTION OF THE APPALACHIAN OROGEN. John M. Bird y John F. Dewey en *Geological Society of America Bulletin,* vol. 81, n.° 4, págs. 1031-1060; abril, 1970.

TECTONICS OF THE WESTERN PIEDMONT AND BLUE RIDGE, SOUTHERN APPALACHIANS: REVIEW AND SPECULATION. Robert D. Hatcher, Jr., en *American Journal of Science,* vol. 278, n.°3, págs. 276-304; marzo, 1978.

THIN-SKINNED TECTONICS IN THE CRYSTALLINE SOUTHERN APPALACHIANS: COCORP SEISMIC-REFLECTION PROFILING OF THE BLUE RIDGE AND PIEDMONT. Frederick A. Cook, Dennis S. Albaugh, Larry D. Brown, Sidney Kaufman, Jack E. Oliver y R. D. Hatcher, Jr., en *Geology,* vol. 7, n.°12, págs. 563-567; diciembre, 1979.

TOMOGRAFIA DE EMISION DE POSITRONES

FATTY ACID UPTAKE AND "METABOLIC IMAGING" OF THE HEART. Milton S. Klein y Burton E. Sobel en *Cardiovascular Clinics,* vol. 10, n.° 2, págs. 165-176; 1979.

QUANTITATIVE IN VIVO AUTORADIOGRAPHY WITH POSITRON-EMISSION TOMOGRAPHY. Marcus E. Raichle en *Brain Research Reviews,* vol. 1, n.° 1, págs. 47-68; julio 1979.

REGIONAL ASSESSMENT OF MYOCARDIAL METABOLIC INTEGRITY IN VIVO BY POSITRON-EMISSION TOMOGRAPHY WITH ^{11}C-LABELED PALMITATE. Michel M. Ter-Pogossian, Milton S. Klein, Joanne Markham, Robert Roberts y Burton E. Sobel en *Circulation,* vol. 61, n.° 2, págs. 242-255; febrero, 1980.

MINERIA DEL PETROLEO

MINING FOR PETROLEUM BY UNDERGROUND METHODS. George S. Rice. U.S. Bureau of Mines Bulletin 351, Departamento de Comercio de los Estados Unidos, 1932.

OIL MINING. A TECHNICAL AND ECONOMIC FEASIBILITY STUDY OF OIL PRODUCTION BY MINING METHODS. Anthony Edey, Bruce A. Kennedy y Leigh A. Readdy en U.S. Bureau of Mines Open File Report 55-79, PB 297 134/AS, Golden Associates, 1978.

LA FLOTABILIDAD DEL NAUTILO

CAMERAL LIQUID IN *NAUTILUS* AND AMMONITES. Peter D. Ward en *Paleobiology,* vol. 5, n.° 1 págs. 40-49; invierno de 1979.

FUNCTION OF CAMERAL WATER IN *NAUTILUS*. Desmond Collins, Peter D. Ward y Gerd E. G. Westermann en *Paleobiology,* vol. 6, n.° 2, págs. 168-172; primavera de 1980.

CAMERAL LIQUID TRANSPORT AND BUOYANCY CONTROL IN THE CHAMBERED NAUTILUS (*NAUTILUS MACROMPHALUS*. Lewis Greenwald, Peter D. Ward y O. E. Greenwald en *Nature,* vol. 286, n.° 5768, págs. 55-56; 3 de julio de 1980.

JUEGOS MATEMATICOS

THEORY OF VOTING. Robin Farquharson. Yale University Press, 1969.

THE PRESIDENTIAL ELECTION GAME. Steven J. Brams. Yale University Press, 1978.

TALLER Y LABORATORIO

VISIBLE LASER TRANSITIONS IN Hg^+. W. E. Bell en *Applied Physics Letters,* vol. 4, n.° 2, págs. 34-35; 15 de enero de 1964.

CREATIVE GLASS BLOWING. James E. Hammesfahr y Clair L. Stong. W. H. Freeman and Company, 1968.

LIGHT AND ITS USES: MAKING AND USING LASERS, HOLOGRAMS, INTERFEROMETERS AND INSTRUMENTS OF DISPERSION. T. D. Walker. W. H. Freeman and Company, 1980.

Indice anual

Este índice contiene los nombres de los autores y artículos aparecidos en Investigación y Ciencia *a partir de enero de 1980. Se incorpora asimismo la relación alfabética de las secciones* Juegos matemáticos *y* Taller y laboratorio.

AUTORES

ARTICULOS

JUEGOS MATEMATICOS

Autor: Martin Gardner

TALLER Y LABORATORIO

Autor: Jearl Walker